Graduate Texts in Mathematics

William Fulton

Algebraic Topology

A First Course

Springer

Graduate Texts in Mathematics 153

Graduate Texts in Mathematics

1 TAKEUTI/ZARING. Introduction to Axiomatic Set Theory. 2nd ed.
2 OXTOBY. Measure and Category. 2nd ed.
3 SCHAEFER. Topological Vector Spaces. 2nd ed.
4 HILTON/STAMMBACH. A Course in Homological Algebra. 2nd ed.
5 MAC LANE. Categories for the Working Mathematician. 2nd ed.
6 HUGHES/PIPER. Projective Planes.
7 SERRE. A Course in Arithmetic.
8 TAKEUTI/ZARING. Axiomatic Set Theory.
9 HUMPHREYS. Introduction to Lie Algebras and Representation Theory.
10 COHEN. A Course in Simple Homotopy Theory.
11 CONWAY. Functions of One Complex Variable I. 2nd ed.
12 BEALS. Advanced Mathematical Analysis.
13 ANDERSON/FULLER. Rings and Categories of Modules. 2nd ed.
14 GOLUBITSKY/GUILLEMIN. Stable Mappings and Their Singularities.
15 BERBERIAN. Lectures in Functional Analysis and Operator Theory.
16 WINTER. The Structure of Fields.
17 ROSENBLATT. Random Processes. 2nd ed.
18 HALMOS. Measure Theory.
19 HALMOS. A Hilbert Space Problem Book. 2nd ed.
20 HUSEMOLLER. Fibre Bundles. 3rd ed.
21 HUMPHREYS. Linear Algebraic Groups.
22 BARNES/MACK. An Algebraic Introduction to Mathematical Logic.
23 GREUB. Linear Algebra. 4th ed.
24 HOLMES. Geometric Functional Analysis and Its Applications.
25 HEWITT/STROMBERG. Real and Abstract Analysis.
26 MANES. Algebraic Theories.
27 KELLEY. General Topology.
28 ZARISKI/SAMUEL. Commutative Algebra. Vol.I.
29 ZARISKI/SAMUEL. Commutative Algebra. Vol.II.
30 JACOBSON. Lectures in Abstract Algebra I. Basic Concepts.
31 JACOBSON. Lectures in Abstract Algebra II. Linear Algebra.
32 JACOBSON. Lectures in Abstract Algebra III. Theory of Fields and Galois Theory.
33 HIRSCH. Differential Topology.
34 SPITZER. Principles of Random Walk. 2nd ed.
35 ALEXANDER/WERMER. Several Complex Variables and Banach Algebras. 3rd ed.
36 KELLEY/NAMIOKA et al. Linear Topological Spaces.
37 MONK. Mathematical Logic.
38 GRAUERT/FRITZSCHE. Several Complex Variables.
39 ARVESON. An Invitation to C^*-Algebras.
40 KEMENY/SNELL/KNAPP. Denumerable Markov Chains. 2nd ed.
41 APOSTOL. Modular Functions and Dirichlet Series in Number Theory. 2nd ed.
42 SERRE. Linear Representations of Finite Groups.
43 GILLMAN/JERISON. Rings of Continuous Functions.
44 KENDIG. Elementary Algebraic Geometry.
45 LOÈVE. Probability Theory I. 4th ed.
46 LOÈVE. Probability Theory II. 4th ed.
47 MOISE. Geometric Topology in Dimensions 2 and 3.
48 SACHS/WU. General Relativity for Mathematicians.
49 GRUENBERG/WEIR. Linear Geometry. 2nd ed.
50 EDWARDS. Fermat's Last Theorem.
51 KLINGENBERG. A Course in Differential Geometry.
52 HARTSHORNE. Algebraic Geometry.
53 MANIN. A Course in Mathematical Logic.
54 GRAVER/WATKINS. Combinatorics with Emphasis on the Theory of Graphs.
55 BROWN/PEARCY. Introduction to Operator Theory I: Elements of Functional Analysis.
56 MASSEY. Algebraic Topology: An Introduction.
57 CROWELL/FOX. Introduction to Knot Theory.
58 KOBLITZ. p-adic Numbers, p-adic Analysis, and Zeta-Functions. 2nd ed.
59 LANG. Cyclotomic Fields.
60 ARNOLD. Mathematical Methods in Classical Mechanics. 2nd ed.
61 WHITEHEAD. Elements of Homotopy

(continued after index)

William Fulton

Algebraic Topology

A First Course

With 137 Illustrations

Springer

William Fulton
Mathematics Department
University of Chicago
Chicago, IL 60637
USA

Mathematics Subject Classifications (1991): 55-01

Library of Congress Cataloging-in-Publication Data
Fulton, William, 1939–
Algebraic topology/William Fulton.
p. cm. — (Graduate texts in mathematics)
Includes bibliographical references and index.
ISBN-13: 978-0-387-94327-5
1. Algebraic topology. I. Title. II. Series.
QA612.F85 1995
514′.2—dc20 94-21786

ISBN-13: 978-0-387-94327-5 e-ISBN: 978-1-4612-4180-5
DOI: 10.1007/978-1-4612-4180-5

(EB)

9 8 7 6 5

springeronline.com

To the memory of my parents

Preface

To the Teacher. This book is designed to introduce a student to some of the important ideas of algebraic topology by emphasizing the relations of these ideas with other areas of mathematics. Rather than choosing one point of view of modern topology (homotopy theory, simplicial complexes, singular theory, axiomatic homology, differential topology, etc.), we concentrate our attention on concrete problems in low dimensions, introducing only as much algebraic machinery as necessary for the problems we meet. This makes it possible to see a wider variety of important features of the subject than is usual in a beginning text. The book is designed for students of mathematics or science who are not aiming to become practicing algebraic topologists—without, we hope, discouraging budding topologists. We also feel that this approach is in better harmony with the historical development of the subject.

What would we like a student to know after a first course in topology (assuming we reject the answer: half of what one would like the student to know after a second course in topology)? Our answers to this have guided the choice of material, which includes: understanding the relation between homology and integration, first on plane domains, later on Riemann surfaces and in higher dimensions; winding numbers and degrees of mappings, fixed-point theorems; applications such as the Jordan curve theorem, invariance of domain; indices of vector fields and Euler characteristics; fundamental groups and covering spaces; the topology of surfaces, including intersection numbers; relations with complex analysis, especially on Riemann sur-

faces; ideas of homology, De Rham cohomology, Čech cohomology, and the relations between them and with fundamental groups; methods of calculation such as the Mayer–Vietoris and Van Kampen theorems; and a taste of the way algebra and "functorial" ideas are used in the subject.

To achieve this variety at an elementary level, we have looked at the first nontrivial instances of most of these notions: the first homology group, the first De Rham group, the first Čech group, etc. In the case of the fundamental group and covering spaces, however, we have bowed to tradition and included the whole story; here the novelty is on the emphasis on coverings arising from group actions, since these are what one is most likely to meet elsewhere in mathematics.

We have tried to do this without assuming a graduate-level knowledge or sophistication. The notes grew from undergraduate courses taught at Brown University and the University of Chicago, where about half the material was covered in one-semester and one-quarter courses. By choosing what parts of the book to cover—and how many of the challenging problems to assign—it should be possible to fashion courses lasting from a quarter to a year, for students with many backgrounds. Although we stress relations with analysis, the analysis we require or develop is certainly not "hard analysis."

We start by studying questions on open sets in the plane that are probably familiar from calculus: When are path integrals independent of path? When are 1-forms exact? (When do vector fields have potential functions?) This leads to the notion of winding number, which we introduce first for differentiable paths, and then for continuous paths. We give a wide variety of applications of winding numbers, both for their own interest and as a sampling of what can be done with a little topology. This can be regarded as a glimpse of the general principle that algebra can be used to distinguish topological features, although the algebra (an integer!) is fairly meager.

We introduce the first De Rham cohomology group of a plane domain, which measures the failure of closed forms to be exact. We use these groups, with the ideas of earlier chapters, to prove the Jordan curve theorem. We also use winding numbers to study the singularities of vector fields. Then 1-chains are introduced as convenient objects to integrate over, and these are used to construct the first homology group. We show that for plane open sets homology, winding numbers, and integrals all measure the same thing; the proof follows ideas of Brouwer, Artin, and Ahlfors, by approximating with grids.

As a first excursion outside the plane, we apply these ideas to sur-

faces, seeing how the global topology of a surface relates to local behavior of vector fields. We also include applications to complex analysis. The ideas used in the proof of the Jordan curve theorem are developed more fully into the Mayer–Vietoris story, which becomes our main tool for calculations of homology and cohomology groups.

Standard facts about covering spaces and fundamental groups are discussed, with emphasis on group actions. We emphasize the construction of coverings by patching together trivial coverings, since these ideas are widely used elsewhere in mathematics (vector bundles, sheaf theory, etc.), and Čech cocycles and cohomology, which are widely used in geometry and algebra; they also allow, following Grothendieck, a very short proof of the Van Kampen theorem. We prove the relation among the fundamental group, the first homology group, the first De Rham cohomology group, and the first Čech cohomology group, and the relation between cohomology classes, differential forms, and the coverings arising from multivalued functions.

We then turn to the study of surfaces, especially compact oriented surfaces. We include the standard classification theorem, and work out the homology and cohomology, including the intersection pairing and duality theorems in this context. This is used to give a brief introduction to Riemann surfaces, emphasizing features that are accessible with little background and have a topological flavor. In particular, we use our knowledge of coverings to construct the Riemann surface of an algebraic curve; this construction is simple enough to be better known than it is. The Riemann–Roch theorem is included, since it epitomizes the way topology can influence analysis. Finally, the last part of the book contains a hint of the directions the subject can go in higher dimensions. Here we do include the construction and basic properties of general singular (cubical) homology theory, and use it for some basic applications. For those familiar with differential forms on manifolds, we include the generalization of De Rham theory and the duality theorems.

The variety of topics treated allows a similar variety of ways to use this book in a course, since many chapters or sections can be skipped without making others inaccessible. The first few chapters could be used to follow or complement a course in point set topology. A course with more algebraic topology could include the chapters on fundamental groups and covering spaces, and some of the chapters on surfaces. It is hoped that, even if a course does not get near the last third of the book, students will be tempted to look there for some idea of where the subject can lead. There is some progression in the level of difficulty, both in the text and the problems. The last few chapters

may be best suited for a graduate course or a year-long undergraduate course for mathematics majors.

We should also point out some of the many topics that are omitted or slighted in this treatment: relative theory, homotopy theory, fibrations, simplicial complexes or simplicial approximation, cell complexes, homology or cohomology with coefficients, and serious homological algebra.

To the Student. Algebraic topology can be thought of as the study of the shapes of geometric objects. It is sometimes referred to in popular accounts as "rubber-sheet geometry." In practice this means we are looking for properties of spaces that are unchanged when one space is deformed into another. "Doughnuts and teacups are topologically the same." One problem of this type goes back to Euler: What relations are there among the numbers of vertices, edges, and faces in a convex polytope, such as a regular solid, in space? Another early manifestation of a topological idea came also from Euler, in the Königsberg bridge problem: When can one trace out a graph without traveling over any edge twice? Both these problems have a feature that characterizes one of the main attractions, as well as the power, of modern algebraic topology—that a global question, depending on the overall shape of a geometric object, can be answered by data that are collected locally. Since these are so appealing—and perhaps to capture your interest while we turn to other topics—they are included as problems with hints at the end of this Preface.

In fact, modern topology grew primarily out of its relation with other subjects, particularly analysis. From this point of view, we are interested in how the shape of a geometric object relates to, or controls, the answers to problems in analysis. Some typical and historically important problems here are:

(i) whether differential forms ω on a region that are closed ($d\omega = 0$) must be exact ($\omega = d\mu$) depends on the topology of the region;
(ii) the behavior of vector fields on a surface depends on the topology of the surface; and
(iii) the behavior of integrals $\int dx/\sqrt{R(x)}$ depends on the topology of the surface $y^2 = R(x)$, here with x and y complex variables.

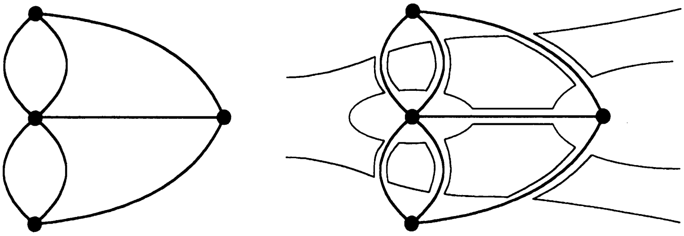
In this book we will begin with the first of these problems, working primarily in open sets in the plane. There is one disadvantage that must be admitted right away: this geometry is certainly flat, and lacks some of the appeal of doughnuts and teacups. Later in the book we

will in fact discuss generalizations to curved spaces like these, but at the start we will stick to the plane, where the analysis is simpler. The topology of open sets in the plane is more interesting than one might think. For example, even the question of the number of connected components can be challenging. The famous Jordan curve theorem, which is one of our goals here, says that the complement of a plane set that is homeomorphic to a circle always has two components—a fact that will probably not surprise you, but whose proof is not so obvious. We will also spend some time on the second problem, which includes the popular problem of whether one can "comb the hair on a billiard ball." We will include some applications to complex analysis, later discussing some of the ideas related to the third problem.

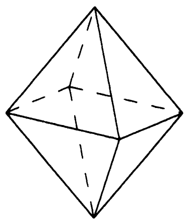
To read this book you need a basic understanding of fundamental notions of the other topology, known as point set topology or general topology. This means that you should know what is meant by words like connected, open, closed, compact, and continuous, and some of the basic facts about them. The notions we need are recalled in Appendix A; if most of this is familiar to you, you should have enough prerequisites. Because of our approach via analysis, you will also need to know some basic facts about calculus, mainly for functions of one or two variables. These calculus facts are set out in Appendix B. In algebra you will need some basic linear algebra, and basic notions about groups, especially abelian groups, which are recalled or proved in Appendix C.

There will be many sorts of *exercises*. Some exercises will be routine applications of or variations on what is done in the text. Those requiring (we estimate) a little more work or ingenuity will be called *problems*. Many will have hints at the end of the book, for you to avoid looking at. There will also be some *projects*, which are things to experiment on, speculate about, and try to develop on your own. For example, one general project can be stated right away: as we go along, try to find analogues in 3-space or n-space for what we do in the plane. (Some of this project is carried out in Part XI.)

Problem 0.1. Suppose X is a graph, which has a finite number of vertices (points) and edges (homeomorphic to a closed interval), with each edge having its endpoints at vertices, and otherwise not intersecting each other. Assume X is connected. When, and how, can you trace out X, traveling along each edge just once? Can you prove your answer?



Exercise 0.2. Let v, e, and f be the number of vertices, edges, and faces on a convex polyhedron. Compute these numbers for the five regular solids, for prisms, and some others. Find a relation among them. Experiment with other polyhedral shapes.



$v = 6$, $e = 12$, $f = 8$

(Note: This problem is "experimental." Proofs are not expected.)

Acknowledgments. I would like to thank the students who came up with good ideas that contributed to notes for the courses, especially K. Ryan, J. Silverman, J. Linhart, J. Trowbridge, and G. Gutman. Thanks also to my colleagues, for answering many of my questions and making useful suggestions, especially A. Collino, J. Harris, R. MacPherson, J.P. May, R. Narasimhan, M. Rothenberg, and S. Weinberger. I am grateful to Chandler Fulton for making the drawings. I would most like to thank those who first inspired me with some of these ideas in courses about three decades ago: H. Federer, J. Milnor, and J. Moore.

William Fulton

Preface to Corrected Edition

I am grateful to J. McClure, J. Buhler, D. Goldberg, and R.B. Burckel for pointing out errors and misprints in the first edition, and for useful suggestions.

Contents

PART VIII
COHOMOLOGY AND HOMOLOGY, III

PART IX
TOPOLOGY OF SURFACES

PART X
RIEMANN SURFACES

PART I

CALCULUS IN THE PLANE

In this first part we will recall some basic facts about differentiable functions, forms, and vector fields, and integration over paths. Much of this should be familiar, although perhaps from a different point of view. At any rate, several of these notions will be needed later, so we take this opportunity to fix the ideas and notation. And of course, we will be looking particularly at the role played by the shapes (topology) of the underlying regions where these things are defined. For the facts that we use, see Appendix B either for precise statements or proofs. Most of this material is included mainly for motivation, and will be developed from a purely topological point of view later; one fact proved in the first chapter—that a closed 1-form on an open rectangle is the differential of a function—will be used later.

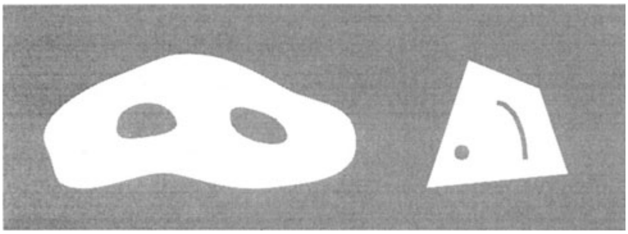
In the second chapter we will see that for any smooth path not passing through the origin, it is possible to define a smooth function that measures how the angle is changing as one moves along the path. This gives us a notion of winding number—how many times a closed path "goes around" the origin. Facts about changing variables in integrals are used to see what happens to integrals and winding numbers when paths are reparametrized and deformed. The third section includes a reinterpretation of the facts from the first chapter in vector field language, and gives a physical interpretation of these ideas to fluid flow. Although we will not use these facts in the book, we will study vector fields later, and it should be useful to have some feeling for them, if you don't already.

CHAPTER 1

Path Integrals

1a. Differential Forms and Path Integrals

In this chapter, U will denote an open set in the plane $\mathbb{R}^2$, for example, the unshaded part of



A *smooth* or $\mathscr{C}^\infty$ *function* on U is a function $f: U \to \mathbb{R}$ such that all partial derivatives of all orders[1] exist and are continuous. In particular, its partial derivatives $\partial f/\partial x$ and $\partial f/\partial y$ are $\mathscr{C}^\infty$ functions on U. Since in this chapter we will only consider $\mathscr{C}^\infty$ functions, we will sometimes just call them functions.

A function f on U is called *locally constant* if every point of U has a neighborhood on which f is constant.

[1] We will never need more than continuous second derivatives, and often much less. The few functions that we actually use, however, will be infinitely differentiable. The extra hypotheses are included so we never have to worry about differentiating any function we meet. The analytically inclined reader may enjoy supplying minimal hypotheses for each assertion.

Exercise 1.1. Prove that a function on an open set U in the plane is locally constant if and only if it is constant on each connected component of U. In other words, defining a locally constant function on U is the same as specifying a constant for each of its connected components.

If f is locally constant, then $\partial f/\partial x = 0$ and $\partial f/\partial y = 0$ (identically, as functions on U), as follows immediately from the definitions of partial derivatives. The converse is also true and only slightly harder:

Proposition 1.2. *If f is a smooth function on U, then f is locally constant if and only if $\partial f/\partial x = 0$ and $\partial f/\partial y = 0$.*

Proof. The point is that, in a rectangular neighborhood of a point of U, the condition $\partial f/\partial x = 0$ means that f is independent of x, i.e., that f is constant along horizontal lines. Likewise $\partial f/\partial y = 0$ means that f is constant along vertical lines, and both conditions make f constant in the rectangle. □

It may not be much, but there is a grain of topology in this:

Corollary 1.3. *The open set U is connected if and only if every smooth function f in U with $\partial f/\partial x = 0$ and $\partial f/\partial y = 0$ is constant.* □

A *differential* 1*-form*, or just a 1*-form*, on U is given by a pair of smooth functions p and q on U. We will usually denote a 1-form by ω, and we will write $\omega = p\,dx + q\,dy$. This can be regarded as just a formal notation, with the dx and dy there merely to indicate what we will do with 1-forms, namely integrate them over paths. The pair of functions (p, q) can also be identified with a vector field on U. For this interpretation, see §2c in Chapter 2.

By a *smooth path* (just called a *path* in this chapter) in U, we mean a mapping $\gamma: [a,b] \to U$ from a bounded interval into U that is continuous on $[a,b]$ and differentiable in the open interval (a,b); in addition, to avoid any trouble at the endpoints, we assume the two component functions of γ can be extended to $\mathscr{C}^\infty$ functions in some neighborhood of $[a,b]$. So $\gamma(t) = (x(t), y(t))$, where x and y are restrictions of smooth functions on an interval[2] $(a - \varepsilon, b + \varepsilon)$, for some

[2] In fact, there are many extensions of these functions to such neighborhoods, but we will never care about values outside the interval $[a,b]$. The assumption is useful to assure that the derivatives of these functions are continuous on the whole closed interval $[a,b]$.

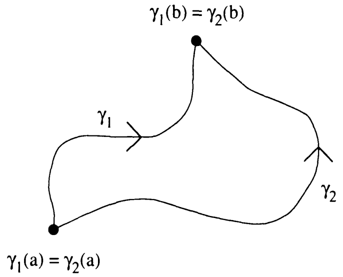
positive number ε. We call $\gamma(a)$ the *initial* point of γ, and $\gamma(b)$ the *final* point; $\gamma(a)$ and $\gamma(b)$ are called the *endpoints*, and we say that γ is a path *from* $\gamma(a)$ *to* $\gamma(b)$.

With $\omega = p\,dx + q\,dy$ as above, and γ a path given by the pair of functions $\gamma(t) = (x(t), y(t))$, the *integral* $\int_\gamma \omega$ *of* ω *along* γ is defined by the formula

$$\int_\gamma \omega \;=\; \int_a^b \left(p(x(t), y(t))\frac{dx}{dt} + q(x(t), y(t))\frac{dy}{dt} \right) dt .$$

Note that the integrand is continuous on $[a, b]$, so the integral exists, as a limit of Riemann sums.

The question we will be concerned with is this: given a 1-form ω on U, when does the integral $\int_\gamma \omega$ depend only on the endpoints $\gamma(a)$ and $\gamma(b)$ of γ, and not on the actual path between them?



Language is usually abused here, saying the integral is "independent of path." This happens whenever there is a "potential function":

Proposition 1.4. *If* $\omega = \partial f/\partial x\, dx + \partial f/\partial y\, dy$, *for some* $\mathscr{C}^\infty$ *function* f *on an open set containing the path* γ, *then*

$$\int_\gamma \omega \;=\; f(\gamma(b)) - f(\gamma(a)) .$$

Proof. Since, by the chain rule,

$$\frac{d}{dt}(f(\gamma(t))) \;=\; \frac{\partial f}{\partial x}(x(t), y(t))\frac{dx}{dt} + \frac{\partial f}{\partial y}(x(t), y(t))\frac{dy}{dt},$$

the integral is

$$\int_\gamma \omega \;=\; \int_a^b \frac{d}{dt}(f(\gamma(t)))\, dt \;=\; f(\gamma(b)) - f(\gamma(a)),$$

the last step by the fundamental theorem of calculus. □

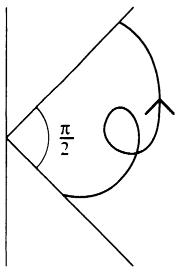
We write $df = \partial f/\partial x\, dx + \partial f/\partial y\, dy$ for this 1-form, and say that ω is the *differential* of f if $\omega = df$.

Exercise 1.5. Show that $df = dg$ on U if and only if $f - g$ is locally constant on U.

For an example, take U to be the right half plane, i.e., the set of points (x, y) with $x > 0$. Consider the function f that measures the angle in polar coordinates, measured counterclockwise from the x-axis. Analytically, $f(x, y) = \tan^{-1}(y/x)$, so

$$\begin{aligned} df &= \frac{1}{1+(y/x)^2}\left(-\frac{y}{x^2}\right)dx + \frac{1}{1+(y/x)^2}\left(\frac{1}{x}\right)dy \\ &= \frac{-y}{x^2+y^2}dx + \frac{x}{x^2+y^2}dy = \frac{-y\,dx + x\,dy}{x^2+y^2}. \end{aligned}$$

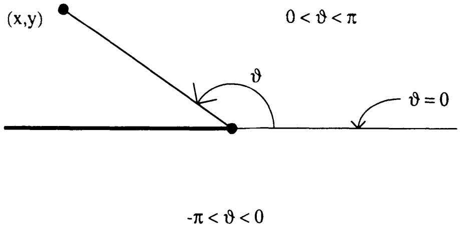
For example, if γ is any path in U from $(1, -1)$ to $(2, 2)$, then $\int_\gamma df = \pi/2$, since that is the change in angle between the two points.



Although the function $f(x, y) = \tan^{-1}(y/x)$, at least as it stands, is not defined where $x = 0$, the expression we found for df makes sense everywhere except at the origin, and is a smooth 1-form on the open set $\mathbb{R}^2 \setminus \{(0, 0)\}$. Let us denote this 1-form by ω_ϑ:

$$\omega_\vartheta = \frac{-y\,dx + x\,dy}{x^2+y^2} \quad \text{on} \quad \mathbb{R}^2 \setminus \{(0,0)\}.$$

In fact, although y/x cannot be extended across the y-axis, the function $\tan^{-1}(y/x)$ can, at least away from the origin. This is clear if we think of it geometrically as the angle in polar coordinates, which can be extended, for example, to the complement of the negative x-axis:



However, there is trouble in trying to extend this angle function to be well defined everywhere on $\mathbb{R}^2 \setminus \{(0,0)\}$. In fact, we can use our last proposition to show that there is no smooth function g on $\mathbb{R}^2 \setminus \{(0,0)\}$ with $dg = \omega_\vartheta$. For example, if $\gamma(t) = (\cos(t), \sin(t))$, $0 \le t \le 2\pi$, is the counterclockwise path around the unit circle, we calculate using the definition of the path integral:

$$\int_\gamma \omega_\vartheta = \int_0^{2\pi} (-\sin(t) \cdot (-\sin(t)) + \cos(t) \cdot \cos(t))\, dt$$
$$= \int_0^{2\pi} 1\, dt = 2\pi .$$

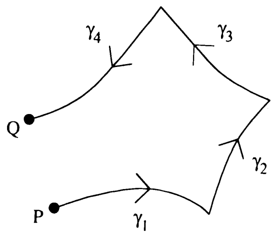
Since $\gamma(0) = \gamma(2\pi)$, it follows from Proposition 1.4 that ω_ϑ cannot be the differential of any function.

Exercise 1.6. On which of the following open sets U is there a smooth function g with $dg = \omega_\vartheta$ on U? Prove your answers. (i) The upper half plane $\{(x, y)\colon y > 0\}$. (ii) The union of the upper half plane and the right half plane. (iii) The left half plane. (iv) The lower half plane. (v) The complement of the negative x-axis. (vi) The annulus $\{(x, y)\colon 1 < x^2 + y^2 < 2\}$. (vii) *Challenge*. The points of the form $(re^t \cos(t), re^t \sin(t))$, $0 < t < 4\pi$, $1/2 < r < 2$.

Exercise 1.7. Is $\omega = (x\,dx + y\,dy)/(x^2 + y^2)^2$ the differential of a function on $\mathbb{R}^2 \setminus \{(0,0)\}$?

1b. When Are Path Integrals Independent of Path?

It will be useful to generalize the notion of smooth path in order to allow integration over a sequence of such paths. Let us define a *segmented path* γ to be a sequence of paths $\gamma_1, \gamma_2, \ldots, \gamma_n$, where each γ_i is a smooth path, and the final point of each γ_i is the initial point of the next γ_{i+1}, for $i = 1, 2, \ldots, n-1$.



We sometimes write $\gamma = \gamma_1 + \ldots + \gamma_n$ for this segmented path. The *initial* point of γ is defined to be the initial point of γ_1, and the *final* point of γ is defined to be the final point of γ_n. The segmented path is *closed* if the final point of γ_n is the initial point of γ_1. If ω is a 1-form on an open set containing (the images of) these paths, we define

$$\int_\gamma \omega = \int_{\gamma_1} \omega + \int_{\gamma_2} \omega + \ldots + \int_{\gamma_n} \omega .$$

If γ is a segmented path in U from P to Q, and $\omega = df$ in U, then it follows from Proposition 1.4 (the "interior endpoints" canceling) that

$$\int_\gamma \omega = f(Q) - f(P).$$

We'll show now that the converse of this is also true:

Proposition 1.8. *Let ω be a 1-form on U. The following are equivalent*: (i) $\int_\gamma \omega = \int_\delta \omega$ *for all segmented paths γ and δ in U with the same initial and final points*; (ii) $\int_\tau \omega = 0$ *for all segmented paths τ in U that are closed*; *and* (iii) $\omega = df$ *for some smooth function f on U.*

Proof. The preceding remark shows that (iii) implies (i). To show that (ii) implies (i), we use the notion of the *inverse* of a path $\sigma\colon [a,b] \to U$, which is the path $\sigma^{-1}\colon [a,b] \to U$ defined by $\sigma^{-1}(t) = \sigma(b + a - t)$; note that the integral of any ω along σ^{-1} is the negative of the integral of ω along σ (cf. Exercise 2.12). Given γ and δ as in (ii), form the closed segmented path τ which is first the sequence of paths making up γ, and then the inverses of the paths that make up δ, but taken in the reverse order. Then $\int_\tau \omega = \int_\gamma \omega - \int_\delta \omega$, from which the fact that (ii) implies (i) follows. That (i) implies (ii) is obvious, by comparing a closed path with a constant path. To show that (i) implies (iii), it is enough to find such a function on each connected component of U, so we can assume U is connected, and hence path-connected (see Appendix A2). Choose and fix an arbitrary

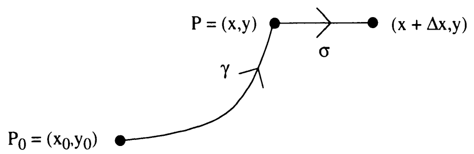
point P_0 in U, and define a function f on U by the formula

$$f(P) = \int_\gamma \omega,$$

where γ is any segmented path from P_0 to P in U. (See Exercise 1.9 below.) By assumption, this is a well-defined function on U. We claim that $\partial f/\partial x = p$ and $\partial f/\partial y = q$, where $\omega = p\,dx + q\,dy$. For the first, we must look at the limit of

$$\frac{f(x+\Delta x, y) - f(x,y)}{\Delta x}$$

as Δx approaches zero, and $P = (x,y)$ is any point in U. To estimate this, let σ be the path from (x,y) to $(x+\Delta x, y)$ given by the formula $\sigma(t) = (x+t, y)$ $0 \le t \le \Delta x$, assuming for the moment that Δx is positive. Let γ be any segmented path from P_0 to P.



Since $\gamma + \sigma$ is a segmented path from P_0 to $(x+\Delta x, y)$,

$$\frac{f(x+\Delta x, y) - f(x,y)}{\Delta x} = \frac{1}{\Delta x}\left(\int_{\gamma+\sigma}\omega - \int_\gamma \omega\right) = \frac{1}{\Delta x}\left(\int_\sigma \omega\right)$$
$$= \frac{1}{\Delta x}\int_0^{\Delta x} p(x+t, y)\,dt.$$

By the mean value theorem, this last expression is equal to $p(x^*, y)$ for some x^* between x and $x + \Delta x$. Letting $\Delta x \to 0$, we have, since p is continuous,

$$\lim_{\Delta x \to 0} \frac{1}{\Delta x}\int_0^{\Delta x} p(x+t, y)\,dt = p(x,y),$$

as required. If Δx is negative, use instead the path $\sigma(t) = (x - t, y)$, $0 \le t \le |\Delta x|$, and the argument is the same, with the modification

$$\frac{1}{\Delta x}\left(\int_\sigma \omega\right) = \frac{-1}{|\Delta x|}\int_0^{|\Delta x|} p(x-t, y)(-1)\,dt$$
$$= \frac{1}{|\Delta x|}\int_0^{|\Delta x|} p(x-t, y)\,dt = p(x^*, y),$$

with $x + \Delta x \leq x^* \leq x$. The proof that $df/\partial y = q$ is similar, interchanging the roles of x and y, and is left as an exercise. □

Exercise 1.9. Show that an open set U in the plane is connected if and only if there is a segmented path between any two points of U. *Challenge.* Can you show that any two points in a connected open set can be connected by an *arc*, i.e., a path that is one-to-one, and whose tangent vector never vanishes?

1c. A Criterion for Exactness

We want a practical criterion to tell if a given 1-form ω is the differential of some function without going to the work of constructing such a function. We shall need another fact from calculus, the equality of mixed partial derivatives: $\partial/\partial x(\partial f/\partial y) = \partial/\partial y(\partial f/\partial x)$. This translates to a simple

Criterion 1.10. *$\omega = p\,dx + q\,dy$ cannot be the differential of a function unless $\partial q/\partial x = \partial p/\partial y$.*

This necessary condition, however, is not always sufficient. For example, if ω_ϑ is the 1-form on $U = \mathbb{R}^2 \setminus \{(0,0)\}$ that we looked at earlier, you can verify easily that ω_ϑ satisfies this condition—either by direct calculation, or by the fact that any point in U has a neighborhood on which the restriction of ω_ϑ is the differential of a function—but we have seen that ω_ϑ cannot be the differential of any function on U. We also saw that the restrictions of ω_ϑ to some simpler open sets, like the right half plane, are the differentials of functions. We will see that it is the topology of U that is controlling this situation.

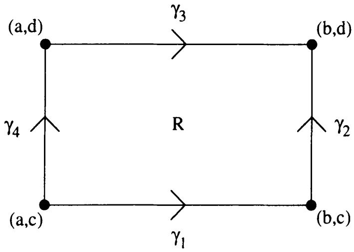
A *2-form* on U is an expression $h\,dx\,dy$, where h is a $\mathscr{C}^\infty$ function on U. Logically, as before, a 2-form can be identified with the function h that defines it, with the "$dx\,dy$" playing only a formal role. The notation indicates that we will use 2-forms for integrating over two-dimensional regions. All we will need is double integrals $\iint_R h\,dx\,dy$ over rectangles $R = [a,b] \times [c,d]$, defined as limits of Riemann sums.

If $\omega = p\,dx + q\,dy$ is a 1-form on U, define $d\omega$ to be the 2-form

$$d\omega = \left(\frac{\partial q}{\partial x} - \frac{\partial p}{\partial y}\right) dx\,dy.$$

So our criterion can be stated: if $\omega = df$, then $d\omega = 0$; or simply that $d(df) = 0$ for all functions f.

Let $R = [a, b] \times [c, d]$ be a closed (bounded) rectangle, and consider the four boundary segments:



In formulas,

$$\gamma_1(t) = (t, c), \quad a \le t \le b; \qquad \gamma_2(t) = (b, t), \quad c \le t \le d;$$
$$\gamma_3(t) = (t, d), \quad a \le t \le b; \qquad \gamma_4(t) = (a, t), \quad c \le t \le d.$$

We will need Green's theorem for a rectangle (see Appendix B). This says that if ω is a 1-form on an open set containing the rectangle R, then

$$\int_{\partial R} \omega = \iint_R d\omega,$$

where

$$\int_{\partial R} \omega = \int_{\gamma_1} \omega + \int_{\gamma_2} \omega - \int_{\gamma_3} \omega - \int_{\gamma_4} \omega.$$

We will need only the following consequence:

Lemma 1.11. *If* $d\omega = 0$, *then* $\int_{\partial R} \omega = 0$, *i.e.*,

$$\int_{\gamma_1} \omega + \int_{\gamma_2} \omega = \int_{\gamma_3} \omega + \int_{\gamma_4} \omega.$$

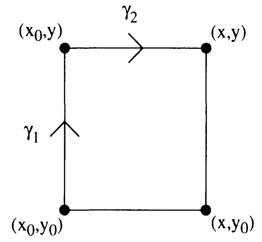
We can apply this to show that, on the plane, or a half plane, or any rectangle, the necessary condition $d\omega = 0$ is actually sufficient for integrals to be path-independent:

Proposition 1.12. *Let U be a product of two open finite or infinite intervals, i.e.*,

$$U = \{(x, y) : a < x < b \quad \text{and} \quad c < y < d\},$$

with $-\infty \le a < b \le \infty$ *and* $-\infty \le c < d \le \infty$. *If* ω *is any* 1*-form on* U *such that* $d\omega = 0$, *then there is a function* f *on* U *with* $\omega = df$.

Proof. Fix a point $P_0 = (x_0, y_0)$ in U. For $P = (x, y)$ in U, let $f(P) = \int_\gamma \omega$, where $\gamma = \gamma_1 + \gamma_2$ is the path shown:



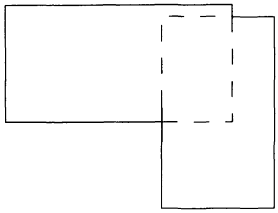
Assume for the moment that $x \ge x_0$ and $y \ge y_0$. The formulas are $\gamma_1(t) = (x_0, y_0 + t)$, $0 \le t \le y - y_0$, and $\gamma_2(t) = (x_0 + t, y)$, $0 \le t \le x - x_0$. The last calculation in the proof of Proposition 1.8 shows that $\partial f/\partial x = p$, where $\omega = p\,dx + q\,dy$. If $y < y_0$, the same is true, replacing $\gamma_1(t)$ by $(x_0, y_0 - t)$, $0 \le t \le y_0 - y$; and similarly if $x < x_0$, replace $\gamma_2(t)$ by $(x_0 - t, y)$, $0 \le t \le x_0 - x$.

Similarly, define a function g by $g(P) = \int_{\gamma^*} \omega$, where γ^* is the path that first goes horizontally from (x_0, y_0) to (x, y_0) and then goes vertically from (x, y_0) to (x, y). The same argument, again left as an exercise, shows that $\partial g/\partial y = q$. Our assumptions on U imply that the closed rectangle with opposite corners at P_0 and P is contained in U, so that Green's theorem can be applied, and it follows from Lemma 1.11 that $f(P)$ is equal to $g(P)$. It follows that $\partial f/\partial x = p$ and $\partial f/\partial y = q$, which means that $df = \omega$. □

Exercise 1.13. Show that the proposition is also true when U is the inside of a disk, i.e., $U = \{(x, y): (x - a)^2 + (y - b)^2 < r^2\}$. Can you prove it when U is any convex region, or any starshaped region? (Convex means that the straight line between any two points in the region is contained in the region, and starshaped means that there is a point P_0 in the region such that for any point P in the region, the straight line from P_0 to P is contained in the region.)

A 1-form ω is called *closed* if $d\omega = 0$, and it is *exact* if $\omega = df$ for some function f. So all exact forms are closed, and the last proposition says that, when U is a rectangle, all closed forms are exact. There are many other regions U for which this is true, besides rectangles and those in the preceding exercise. For example, if U is the union

of any two rectangles, each as in the proposition, then any closed 1-form ω in U is exact.

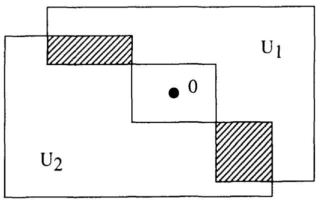


To see this, let U_1 and U_2 be the two open rectangles whose union is U. By the proposition, there are functions f_1 in U_1 and f_2 in U_2 such that $df_1 = \omega$ on U_1 and $df_2 = \omega$ on U_2. Since $U_1 \cap U_2$ is connected, and since $d(f_2 - f_1) = \omega - \omega = 0$ on $U_1 \cap U_2$, it follows that $f_2 - f_1$ is a constant function on $U_1 \cap U_2$. If we replace f_2 by $f_2 - c$, where c is this constant, we can assume f_1 and f_2 agree on $U_1 \cap U_2$. This means that there is a function f on $U = U_1 \cup U_2$ whose restriction to U_1 is f_1 and whose restriction to U_2 is f_2. Moreover, $df = \omega$ on all of U, since this condition is a local condition, to be verified at each point of U; and every point is either in U_1 or U_2, where we know $df_1 = \omega$ and $df_2 = \omega$.

In fact, this argument proves:

Lemma 1.14. *Suppose U_1 and U_2 are open sets, and $U_1 \cap U_2$ is connected. Let $U = U_1 \cup U_2$, and let ω be a 1-form on U. If the restrictions of ω to U_1 and U_2 are both exact, then ω is exact on U.* □

The connectedness of $U_1 \cap U_2$ is crucial for this. For example, suppose U_1 and U_2 are the regions indicated, with the shaded overlap, and the origin outside in the middle.



Let ω be the restriction of ω_ϑ to U, then ω_ϑ is the differential of a function on each of the regions, but not on their union, as can be seen by integrating ω along a path around the origin in U.

Exercise 1.15. Show that if U is a union of open sets $U_1, \ldots, U_n$,

and ω is a 1-form on U such that the restriction of ω to each U_i is exact, and $(U_1 \cup U_2 \cup \ldots \cup U_i) \cap U_{i+1}$ is connected for $1 \le i \le n-1$, then ω is exact on U.

If ω is a closed 1-form in an open set U, although ω may not be exact on all of U, it follows from Proposition 1.12 that it is always locally exact. That is, any point in U has a neighborhood (say rectangular), so that the restriction of ω to this neighborhood is the differential of a function. This can be used to calculate path integrals, by cutting the path into pieces, on each of which Proposition 1.4 can be applied:

Proposition 1.16. *If ω is a closed 1-form on U, and γ: $[a, b] \to U$ is a smooth path, then there is a subdivision $a = t_0 < t_1 < \ldots < t_n = b$ and a collection of open subsets $U_1, \ldots, U_n$ of U so that γ maps $[t_{i-1}, t_i]$ into U_i, and the restriction of ω to U_i is the differential of a function f_i. Let $P_i = \gamma(t_i)$. Then, for any such choices,*

$$\int_\gamma \omega = (f_1(P_1) - f_1(P_0)) + (f_2(P_2) - (f_2(P_1)) + \ldots + (f_n(P_n) - f_n(P_{n-1})).$$

Proof. For each point P in $\gamma([a, b])$, choose a neighborhood U_P of P on which the restriction of ω is exact. The open sets $\gamma^{-1}(U_P)$ form an open covering of the compact interval $[a, b]$, so a finite number of them cover the interval. From this it is not hard to construct the subdivision. One quick way to do it is to use the Lebesgue covering lemma (see §A4 of Appendix A), which guarantees that, if the subdivision is small enough, each subinterval will be mapped into one of the neighborhoods U_P. Having fixed such a subdivision, choose one of these open sets containing the image of $[t_{i-1}, t_i]$, call it U_i, and choose a function f_i on U_i with $df_i = \omega$ on U_i. Let γ_i: $[t_{i-1}, t_i] \to U_i$ be the restriction of γ to $[t_{i-1}, t_i]$. Then

$$\int_\gamma \omega = \int_{\gamma_1} \omega + \int_{\gamma_2} \omega + \ldots + \int_{\gamma_n} \omega$$
$$= \sum_{i=1}^{n} (f_i(\gamma_i(t_i)) - f_i(\gamma_i(t_{i-1}))) = \sum_{i=1}^{n} (f_i(P_i) - f_i(P_{i-1})). \qquad \square$$

The following two lemmas will be used in Part III to prove the Jordan curve theorem. They are special cases of general theorems to be proved in Chapter 9, but we can prove them directly with the

methods of this section. For any positive number r, let $\gamma_{P,r}$ be the counterclockwise circle of radius r about P:

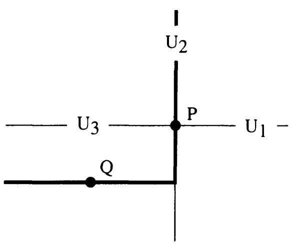
$$\gamma_{P,r}(t) = P + r(\cos(2\pi t), \sin(2\pi t)), \qquad 0 \leq t \leq 1.$$

Lemma 1.17. *Suppose $U = \mathbb{R}^2 \setminus \{P\}$, and let $r > 0$. If ω is a closed 1-form on U such that $\int_{\gamma_{P,r}} \omega = 0$, then ω is exact.*

Proof. Let U_1, U_2, U_3, and U_4 be the half planes to the right of P, above P, to the left of P, and below P. By Proposition 1.12, there are functions f_i on U_i, unique up to the addition of constants, with $df_i = \omega$ on U_i. By adjusting the constants, we may assume $f_2 = f_1$ on $U_1 \cap U_2$, and $f_3 = f_2$ on $U_2 \cap U_3$, and $f_4 = f_3$ on $U_3 \cap U_4$. Then $f_4 = f_1 + c$ on $U_4 \cap U_1$ for some constant c. By Proposition 1.16, $\int_{\gamma_{P,r}} \omega = c$. The hypothesis implies that $c = 0$, which means exactly that the four functions f_i agree on overlaps, so define a function f on U such that $df = \omega$. □

Lemma 1.18. *Suppose $U = \mathbb{R}^2 \setminus \{P, Q\}$, and let $0 < r < |P - Q|$. If ω is a closed 1-form on U such that $\int_{\gamma_{P,r}} \omega = 0$ and $\int_{\gamma_{Q,r}} \omega = 0$, then ω is exact.*

Proof. The proof is similar. Suppose for definiteness that Q is located to the southwest of P. Let U_1 be the half plane to the right of P, and choose f_1 on U_1 so that $\omega = df_1$ on U_1. Let U_2 be the half plane above P, and choose f_2 on U_2 so that $\omega = df_2$ on U_2, and $f_2 = f_1$ on $U_1 \cap U_2$. Let U_3 be the quarter plane to the left of P and above Q, and choose f_3 on U_3 so that $\omega = df_3$ on U_3, and $f_3 = f_2$ on $U_2 \cap U_3$.



Let U_4 be the quarter plane to the right of Q and below P, and choose f_4 on U_4 so that $\omega = df_4$ on U_4, and $f_4 = f_3$ on $U_3 \cap U_4$. Then f_4 and f_1 differ by a constant on $U_4 \cap U_1$, and the hypothesis that $\int_{\gamma_{P,r}} \omega = 0$ implies as in the preceding lemma that this constant is 0. Let U_5 be the half plane to the left of Q, and choose f_5 on U_5 such that $\omega = df_5$ on U_5 and $f_5 = f_3$ on $U_3 \cap U_5$. Let U_6 be the half plane

below Q, and choose f_6 on U_6 such that $\omega = df_6$ on U_6 and $f_6 = f_5$ on $U_5 \cap U_6$. The hypothesis that $\int_{\gamma_{Q,r}} \omega = 0$ implies that $f_6 = f_4$ on $U_4 \cap U_6$. The functions f_i on U_i agree on overlaps, defining a function f on U such that $\omega = df$, which completes the proof in this case.

The proof when Q is located southeast of P, or the special cases when P and Q are on a horizontal or vertical line, are similar and left to the reader. □

Problem 1.19. Generalize the preceding lemmas from one or two to n points.

For any point $P = (x_0, y_0)$, define the 1-form $\omega_{P,\vartheta}$ on $\mathbb{R}^2 \setminus \{P\}$ by the formula

$$\omega_{P,\vartheta} = \frac{-(y - y_0)\,dx + (x - x_0)dy}{(x - x_0)^2 + (y - y_0)^2}.$$

Problem 1.20. For any two points P and Q, show that the 1-form $\omega = \omega_{P,\vartheta} - \omega_{Q,\vartheta}$ is exact on $\mathbb{R}^2 \setminus L$, where L is the line segment from P to Q. *Challenge.* Find a function whose differential is ω.

CHAPTER 2

Angles and Deformations

2a. Angle Functions and Winding Numbers

Any point in the plane can be expressed in polar coordinates, i.e., it can be written in the form $(r\cos(\vartheta), r\sin(\vartheta))$ for some $r \geq 0$ and some real number ϑ. The radius r is unique, being the distance from the origin, or the square root of the sums of the squares of the Cartesian coordinates. At the origin, $r = 0$, and ϑ can be any number. We often denote the origin simply by 0 instead of $(0, 0)$. Except for the origin, the angle ϑ is determined only up to adding integral multiples of 2π. We call any of these numbers *an angle* for the point.

Suppose γ: $[a, b] \to \mathbb{R}^2 \setminus \{0\}$ is a $\mathscr{C}^\infty$ path to the complement of the origin, given in Cartesian coordinates by $\gamma(t) = (x(t), y(t))$. We want to describe this in polar coordinates, that is, to find $\mathscr{C}^\infty$ functions $r(t)$ and $\vartheta(t)$ so that

$$(2.1) \quad \gamma(t) = r(t)(\cos(\vartheta(t)), \sin(\vartheta(t))) = (r(t)\cos(\vartheta(t)), r(t)\sin(\vartheta(t)))$$

for all $a \leq t \leq b$. There is no problem with the function r: it is the distance from the origin, defined by

$$r(t) = \|\gamma(t)\| = \sqrt{x(t)^2 + y(t)^2}.$$

The angle function is not so simple. At any time t, there are many possible angles to choose, all differing by multiples of 2π. If we choose them say all lying in the interval $(-\pi, \pi]$ they would be unique, but then would not vary continuously if the point crosses the negative x-axis.

The initial angle can be chosen arbitrarily. In other words, choose any number ϑ_a so that

$$\gamma(a) = (x(a), y(a)) = (r(a)\cos(\vartheta_a), r(a)\sin(\vartheta_a)).$$

We will require that our function $\vartheta(t)$ satisfy the initial condition $\vartheta(a) = \vartheta_a$. Motivated by the discussion in §1b, we have a candidate for the derivative $\vartheta'(t)$, which is the rate of change of angle: it should be

$$\vartheta'(t) = \frac{-y(t)x'(t) + x(t)y'(t)}{x(t)^2 + y(t)^2} = \frac{-y(t)x'(t) + x(t)y'(t)}{r(t)^2}.$$

So we can define $\vartheta(t)$ to be the unique function with this initial condition and this derivative. That is, we define $\vartheta(t)$ by

$$\vartheta(t) = \vartheta_a + \int_a^t \frac{-y(\tau)x'(\tau) + x(\tau)y'(\tau)}{r(\tau)^2}\, d\tau.$$

Proposition 2.2. *With these definitions, the functions $r(t)$ and $\vartheta(t)$ are $\mathscr{C}^\infty$ functions, and equation* (2.1) *is satisfied for all t in* $[a, b]$.

Proof. The function $r(t)$ is $\mathscr{C}^\infty$ since it is a composite of $\mathscr{C}^\infty$ functions, and $\vartheta(t)$ is $\mathscr{C}^\infty$ since it is the integral of a $\mathscr{C}^\infty$ function. Let

$$u(t) = (\cos(\vartheta(t)), \sin(\vartheta(t))) \quad \text{and} \quad v(t) = (-\sin(\vartheta(t)), \cos(\vartheta(t))).$$

These are perpendicular unit vectors for all t. We want to show that

$$\frac{1}{r(t)}\gamma(t) = u(t)$$

for all t. We are assuming that they are equal for $t = a$. It suffices to show that both sides of this equation have the same dot product with vectors $u(t)$ and $v(t)$, since the difference $(1/r(t))\gamma(t) - u(t)$ would then be perpendicular to two independent vectors, and so would be zero. For the right side of the equation $u(t)$, these dot products are identically 1 and 0, respectively, so it suffices to prove that the same is true for the left side. Since we know the equality of the vectors for $t = a$, it suffices to prove that the derivatives of these dot products

vanish. We are therefore reduced to showing that

$$\frac{d}{dt}\left(\frac{1}{r(t)}\gamma(t)\cdot u(t)\right) \equiv 0 \quad \text{and} \quad \frac{d}{dt}\left(\frac{1}{r(t)}\gamma(t)\cdot v(t)\right) \equiv 0.$$

These are simple verifications, using the usual rules of calculus. Omitting the variable t, and writing $\dot{r}$ in place of $r'(t)$, etc., the first of these derivatives is

$$\frac{r\dot{x}-x\dot{r}}{r^2}\cos(\vartheta) + \frac{r\dot{y}-y\dot{r}}{r^2}\sin(\vartheta) + \frac{x}{r}\dot{\vartheta}(-\sin(\vartheta)) + \frac{y}{r}\dot{\vartheta}\,(\cos(\vartheta))$$

$$= \frac{1}{r^3}\left(\cos(\vartheta)[r^2\dot{x} - xr\dot{r} + yr^2\dot{\vartheta}] + \sin(\vartheta)[r^2\dot{y} - yr\dot{r} - xr^2\dot{\vartheta}]\right).$$

Using the identities $r\dot{r} = x\dot{x} + y\dot{y}$ and $r^2\dot{\vartheta} = -y\dot{x} + x\dot{y}$, one sees that the terms in brackets vanish. The proof that the other derivative vanishes is similar and left as an exercise. □

Exercise 2.3. (a) Show that this function $\vartheta(t)$ is unique, and that in fact it is the only continuous function with $\vartheta(a) = \vartheta_a$ such that (2.1) holds. (b) Show that if ϑ_a is replaced by $\vartheta_a + 2\pi n$ for some n, then the corresponding angle function is $\vartheta(t) + 2\pi n$.

Using this angle function, we can define the (total signed) *change in angle* of the path γ to be $\vartheta(b) - \vartheta(a)$. Note by the preceding exercise that this is independent of the choice of initial angle. Equivalently, this change in angle is

$$\int_a^b \frac{-y(t)x'(t) + x(t)y'(t)}{r(t)^2}\,dt = \int_\gamma \frac{-y\,dx + x\,dy}{x^2+y^2} = \int_\gamma \omega_\vartheta.$$

If a segmented path does not pass through the origin, i.e., it is a path in $\mathbb{R}^2 \setminus \{0\}$, we want to define its *winding number*, which should be the "net" number of times γ goes around the origin, counting the counterclockwise motion as positive, and the clockwise motion as negative. In other words, it is the total signed change in angle, divided by 2π. We will denote it by $W(\gamma, 0)$, the "0" indicating that we are going around the origin. What we have just done shows what the definition should be. We define the *winding number of* γ *around*

the origin by

$$W(\gamma, 0) = \frac{1}{2\pi}\int_{\gamma} \omega_{\vartheta} = \frac{1}{2\pi}\int_{\gamma} \frac{-y\,dx + x\,dy}{x^2 + y^2}.$$

Proposition 2.4. *For any closed segmented path that does not pass through the origin, the winding number of the path around the origin is an integer.*

Proof. The formula given before the proposition defines a number $W(\gamma, 0)$ for any segmented path γ that does not pass through the origin. Suppose γ starts at P and goes to Q, and we choose an angle ϑ_P for P. We claim that $\vartheta_P + 2\pi W(\gamma, 0)$ is an angle for Q. This will prove the result, for when $P = Q$ it says that ϑ_P and $\vartheta_P + 2\pi W(\gamma, 0)$ define the same angle, so they must differ by an integer multiple of 2π.

It suffices to prove the claim for a smooth path, since if γ is a sum of smooth paths, the assertion for each of them implies it for the sum. But when $\gamma\colon [a, b] \to \mathbb{R}^2 \setminus \{0\}$ is smooth, $\vartheta_P + 2\pi W(\gamma, 0)$ is just the value of our angle function $\vartheta(t)$ at $t = b$, starting with $\vartheta(a) = \vartheta_P$, so the result follows from Proposition 2.2. □

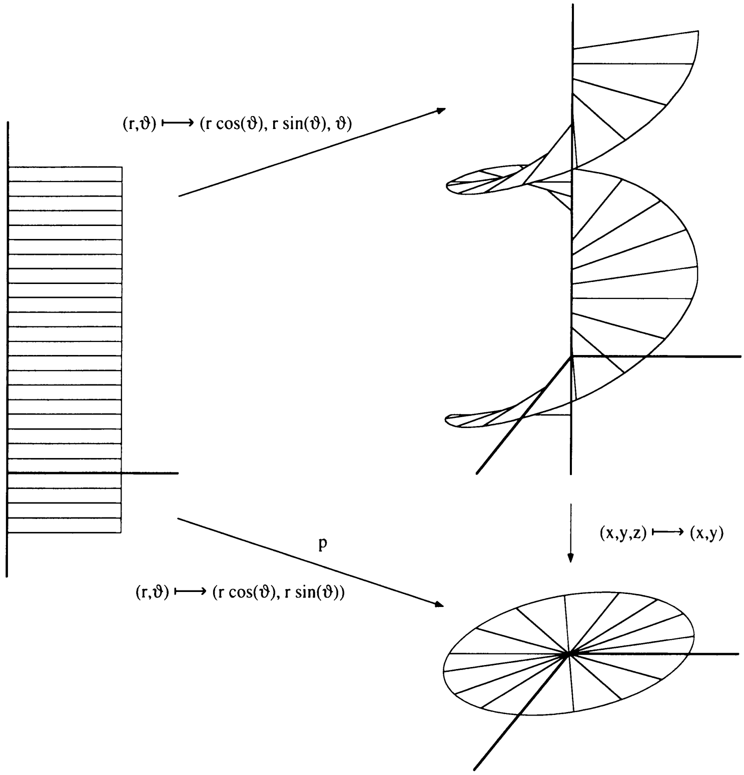
Exercise 2.5. Use Proposition 1.16 to give another proof of this proposition.

Exercise 2.6. Let $\gamma(t) = (k\cos(nt), k\sin(nt))$, $0 \le t \le 2\pi$, where k is a positive number and n is an integer. Show that $W(\gamma, 0) = n$.

In earlier centuries, before modern rigor required functions to be single-valued, the 1-form ω_ϑ would have been written as the differential $d\vartheta$ of the multivalued function ϑ. (In fact, this is still a common and useful notation for this 1-form, as long as one realizes that ϑ is not a function, so that Proposition 1.4 is not contradicted!) The graph of this multivalued function can be visualized in 3-space, as the locus of points (x, y, z) of the form $(r\cos(\vartheta), r\sin(\vartheta), \vartheta)$ for some $r > 0$ and real number ϑ. This is closely related to the *polar coordinate* mapping

$$p\colon \{(r, \vartheta)\colon r > 0\} \to \mathbb{R}^2 \setminus \{0\}, \qquad (r, \vartheta) \mapsto (r\cos(\vartheta), r\sin(\vartheta)).$$

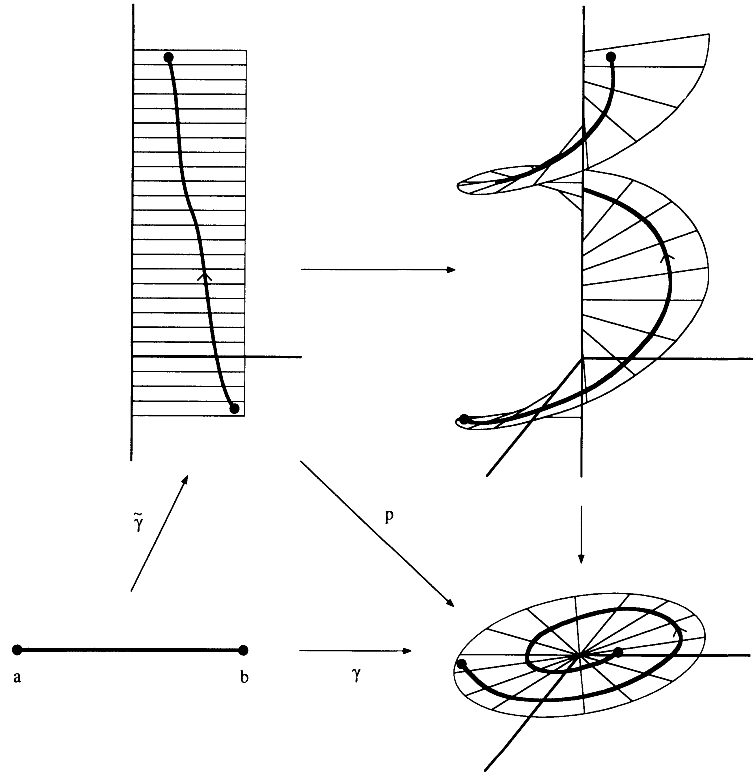
Picturing these together:



This illustrates the fact that, although the distance from the origin is a continuous function on $\mathbb{R}^2 \setminus \{0\}$, the counterclockwise angle from the x-axis cannot be defined continuously. What we have proved, however, is that one can define such an angle continuously along a curve. Proposition 2.2 (with Exercise 2.3) can be expressed geometrically as follows:

Corollary 2.7. *Let* γ: $[a, b] \to \mathbb{R}^2 \setminus \{0\}$ *be a smooth path with starting point* $P_a = (r_a cos(\vartheta_a), r_a sin(\vartheta_a))$. *Then there is a unique smooth path* $\tilde{\gamma}$: $[a, b] \to R$, *where* $R = \{(r, \vartheta): r > 0\}$ *is the right half plane, with starting point* (r_a, ϑ_a), *and with* $p \circ \tilde{\gamma} = \gamma$. □

We say that the path $\tilde{\gamma}$ is the *lifting* of the path γ with starting point (r_a, ϑ_a). On the picture,



Problem 2.8. Show that the locus of points $(r\cos(\vartheta), r\sin(\vartheta), \vartheta)$ is one connected component of the surface in the complement of the z-axis in $\mathbb{R}^3$ defined by the equation $y \cdot \cos(z) = x \cdot \sin(z)$.

Exercise 2.9. Show that every point in $\mathbb{R}^2 \setminus \{0\}$ has a neighborhood V such that $p^{-1}(V)$ is a disjoint union of open sets V_i, each of which is mapped homeomorphically (in fact, diffeomorphically) by p onto V.

This exercise verifies that p is a "covering map," a class of maps we will study in Chapter 11. Corollary 2.7 is a special case of a general theorem about covering maps.

Winding numbers can be described around any point $P = (x_0, y_0)$ in place of the origin. One can do this either by translating everything, or directly by using the form

$$\omega_{P,\vartheta} = \frac{-(y - y_0)\,dx + (x - x_0)dy}{(x - x_0)^2 + (y - y_0)^2}.$$

Exercise 2.10. For any closed segmented path $\gamma\colon [a,b] \to \mathbb{R}^2 \setminus \{P\}$, define the winding number $W(\gamma,P)$ of γ around P by the formula

$$W(\gamma,P) = \frac{1}{2\pi}\int_\gamma \omega_{P,\vartheta}.$$

Generalize the assertions of this section to these winding numbers.

2b. Reparametrizing and Deforming Paths

Path integrals do not depend on the parametrization of the path, in the following sense. Suppose ω is a 1-form on an open set U in the plane, and $\gamma\colon [a,b] \to U$ is a smooth path, and suppose $\varphi\colon [a',b'] \to [a,b]$ is a $\mathscr{C}^\infty$ function (as usual, extending to a neighborhood of $[a',b']$) that maps a' to a and b' to b. The path $\gamma\circ\varphi$ is called a *reparametrization* of γ.

Lemma 2.11. *With these assumptions,*

$$\int_{\gamma\circ\varphi}\omega = \int_\gamma \omega.$$

Proof. Write $\omega = p\,dx + q\,dy$, and $\gamma(t) = (x(t), y(t))$, and calculate, using the chain rule and change of variables formulas:

$$\begin{aligned}\int_{\gamma\circ\varphi}\omega &= \int_{a'}^{b'} [p(x(\varphi(s)), y(\varphi(s)))\cdot(x\circ\varphi)'(s) \\ &\qquad\qquad + q(x(\varphi(s)), y(\varphi(s)))\cdot(y\circ\varphi)'(s)]\,ds \\ &= \int_{a'}^{b'} [p(x(\varphi(s)), y(\varphi(s)))\cdot x'(\varphi(s)) \\ &\qquad\qquad + q(x(\varphi(s)), y(\varphi(s)))\cdot y'(\varphi(s))]\cdot(\varphi)'(s)\,ds \\ &= \int_a^b [p(x(t),y(t))\cdot x'(t) + q(x(t),y(t))\cdot y'(t)]dt = \int_\gamma\omega. \qquad \square\end{aligned}$$

Exercise 2.12. If $\gamma\colon [a,b]\to U$ is a path in U, let $\gamma^{-1}\colon [a,b]\to U$ be the same path traveled backward: $\gamma^{-1}(t) = \gamma(a+b-t)$. Show that

$$\int_{\gamma^{-1}}\omega = -\int_\gamma\omega.$$

Show more generally that if $\varphi\colon [a',b'] \to [a,b]$ is any $\mathscr{C}^\infty$ function that maps a' to b and b' to a, then $\int_{\gamma\circ\varphi}\omega = -\int_\gamma\omega$.

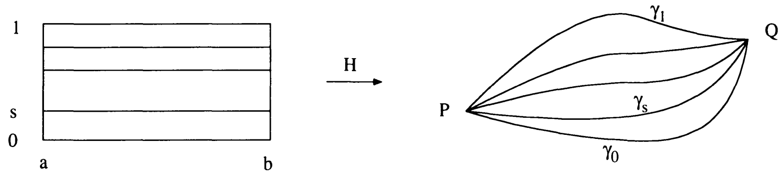
The following problem shows how one could avoid dealing with segmented paths: by turning around the corners *very* slowly!

Problem 2.13. (a) Construct a $\mathscr{C}^\infty$ mapping φ from $[0, 1]$ to $[a, b]$ taking 0 to a and 1 to b, such that all derivatives of φ vanish at 0 and at 1. (b) Use this to show that any smooth path can be reparametrized to a path with the same endpoints but such that all derivatives vanish at the endpoints. (c) If γ_1: $[a, b] \to U$ and γ_2: $[b, c] \to U$ are $\mathscr{C}^\infty$ paths with $\gamma_1(b) = \gamma_2(b)$, and all derivatives of γ_1 and γ_2 vanish at b, verify that the path γ: $[a, c] \to U$ that agrees with γ_1 on $[a, b]$ and γ_2 on $[b, c]$ is a $\mathscr{C}^\infty$ path. (d) If $\gamma = \gamma_1 + \ldots + \gamma_n$ is a segmented path, show that there is a $\mathscr{C}^\infty$ path γ^*: $[0, n] \to U$, so that the restriction of γ^* to $[k-1, k]$ is a reparametrization of γ_k for each k.

Problem 2.14. Given a 1-form ω on an open set U, show that the following are equivalent: (i) there is a function f on U with $df = \omega$; (ii) $\int_\gamma \omega = \int_\delta \omega$ whenever γ and δ are $\mathscr{C}^\infty$ paths in U with the same initial and final points; and (iii) $\int_\gamma \omega = 0$ whenever γ is a $\mathscr{C}^\infty$ closed path in U (i.e., with the initial point equal to the final point).

Problem 2.15. Given a 1-form ω on an open set U, show that the following are equivalent: (i) $d\omega = 0$; (ii) $\int_{\partial R} \omega = 0$ for all closed rectangles R contained in U; and (iii) every point in U has a neighborhood such that $\int_{\partial R} \omega = 0$ for all closed rectangles R contained in the neighborhood. Is the same true if rectangles are replaced by disks?

Next we turn to the question of what happens when the path is moved, or deformed. We consider first the case of a deformation of paths with fixed endpoints. This will be given by a family of paths γ_s, for simplicity all defined on the same interval $[a, b]$, with the parameter s varying in another interval which we take to be the unit interval $[0, 1]$. We will assume this is a smooth family, in the sense that the coordinates of the point $\gamma_s(t)$ are smooth functions of both s and t. This means that we are given a mapping H from $[a, b] \times [0, 1]$ to U, which we assume is $\mathscr{C}^\infty$ (this, as usual, means that the two coordinate functions can be extended to be infinitely differentiable functions on some open neighborhood of the rectangle). We assume that $H(a, s) = P$ and $H(b, s) = Q$ for all $0 \leq s \leq 1$. Set $\gamma_s(t) = H(t, s)$, so each γ_s is a path in U from P to Q.



We call H a *smooth homotopy* from the path γ_0 to the path γ_1. We say that two paths from an interval $[a, b]$ to U, with the same initial and final points, are *smoothly homotopic* in U if there is such a homotopy from one to the other.

Proposition 2.16. *If γ and δ are smoothly homotopic paths from P to Q in an open set U, and ω is a closed 1-form in U, then*

$$\int_\gamma \omega = \int_\delta \omega .$$

Proof. First we sketch a proof using only ideas from calculus. Let V be a neighborhood of the rectangle $R = [a, b] \times [0, 1]$ mapped into U by an extension of H, and let $x(t, s)$ and $y(t, s)$ be the coordinate functions of this mapping from V to U. Define a "pull-back" form ω^* on V by the formula

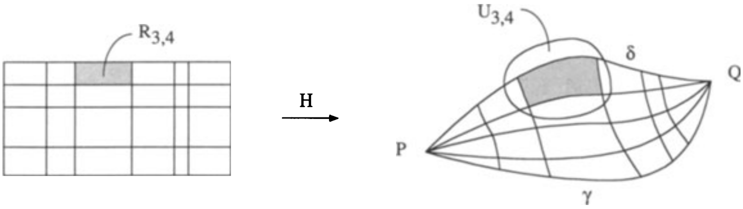
$$\omega^* = \left(p(x(t,s), y(t,s))\frac{\partial x}{\partial t} + q(x(t,s), y(t,s))\frac{\partial y}{\partial t}\right) dt$$
$$+ \left(p(x(t,s), y(t,s))\frac{\partial x}{\partial s} + q(x(t,s), y(t,s))\frac{\partial y}{\partial s}\right) ds .$$

A little calculation, left to you, shows that $d\omega^* = 0$ on V. By Green's theorem, we therefore know that the integral of ω^* around the boundary of R must be zero. Simpler calculations show that the integral of ω^* along the bottom of the rectangle is $\int_\gamma \omega$, that along the top is $\int_\delta \omega$, and that the integrals along the two sides are zero. Putting this all together gives the required equality.

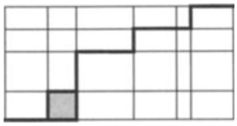
Here is another proof, more topological in flavor. Each point in the image $H(R)$ of the rectangle has a neighborhood on which ω is exact. Applying the Lebesgue lemma, it follows that, if we subdivide the rectangle small enough, by choosing

$$a = t_0 < t_1 < \ldots < t_n = b \quad \text{and} \quad 0 = s_0 < s_1 < \ldots < s_m = 1 ,$$

then each subrectangle $R_{i,j} = [t_{i-1}, t_i] \times [s_{j-1}, s_j]$ is mapped by H into an open set $U_{i,j}$ on which ω is the differential of some function $f_{i,j}$.



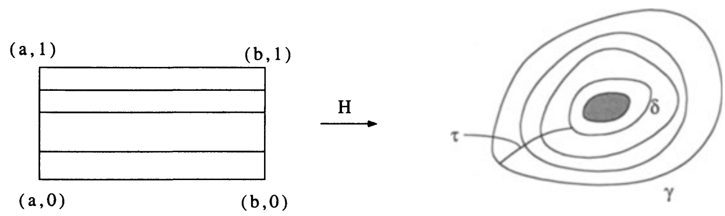
Since ω is the differential of a function on $U_{i,j}$, the integral of $\omega \circ H$ around the boundary of $R_{i,j}$ is zero, i.e., the integral along the bottom and right side of $R_{i,j}$ is the same as the integral along the left side and top. Now $\int_\gamma \omega$ is the integral of ω along the bottom and right side of the original rectangle, and one can successively replace integrals over the bottom and right sides by integrals over the left and top sides, of each of the small rectangles, until one has the integral over the left and top sides of the whole rectangle, which is $\int_\delta \omega$.



More succinctly, integrate ω over all the boundaries of the small rectangles. Each inside edge is integrated over twice, once in each direction, which leaves the integrals over the outside, giving

$$0 = \sum_{i,j} \int_{\partial R_{i,j}} \omega = \int_{\partial R} \omega = \int_\gamma \omega - \int_\delta \omega. \qquad \square$$

There is another kind of deformation or homotopy that is also important. This is a deformation of closed paths, through a family of closed paths, but allowing the endpoints to vary. It is given by a $\mathscr{C}^\infty$ mapping H from a rectangle $[a,b] \times [0,1]$ into U, with the property that $H(a,s) = H(b,s)$ for every s in $[0,1]$. Each of the paths γ_s given by $\gamma_s(t) = H(t,s)$ is a closed path, starting and ending at the point $\tau(s) = H(a,s) = H(b,s)$. These endpoints are allowed to vary along the path τ.



We call H a *smooth homotopy* from the path $\gamma = \gamma_0$ to the path $\delta = \gamma_1$, and we say that two closed paths γ and δ are *smoothly homotopic* if there is such a smooth homotopy between them.

Proposition 2.17. *If γ and δ are smoothly homotopic closed paths in an open set U, and ω is a closed 1-form in U, then*

$$\int_\gamma \omega = \int_\delta \omega .$$

Proof. Either of the proofs of the last proposition works for this one. The only point to notice is that the integrals of ω over the two sides of the rectangle may not be zero, but since the two paths given by H on these two sides are the same (namely τ), these integrals $\int_\tau \omega$ cancel. The details are left as an exercise. □

When U is the complement of a point P, and $\omega = (1/2\pi)\omega_{P,\vartheta}$, this specializes to the important:

Corollary 2.18. *If γ and δ are smoothly homotopic closed paths in $\mathbb{R}^2 \setminus \{P\}$, then $W(\gamma, P) = W(\delta, P)$.*

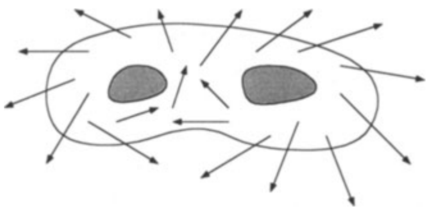
Of course, it is crucial for this that the homotopy stays in the complement of the point P!

Problem 2.19. Prove that being smoothly homotopic is an equivalence relation.

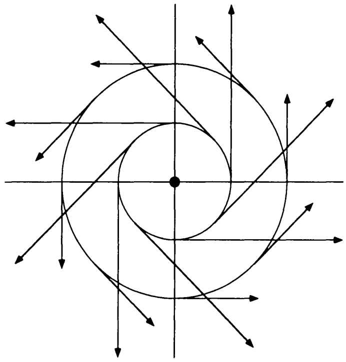
2c. Vector Fields and Fluid Flow

We defined a 1-form on an open set U in the plane to be a pair of smooth functions p and q on U. Such a pair of functions can also be identified with a *vector field* on U, which assigns to each point (x, y) in U the vector

$$V(x, y) = p(x, y)\mathbf{i} + q(x, y)\mathbf{j} .$$



The vector field corresponding to the 1-form ω_ϑ is perpendicular to the position vector, pointing in a counterclockwise direction, with length that is the inverse of the distance to the origin:



The path integral

$$\int_\gamma p\,dx + q\,dy = \int_a^b \left(p(x(t), y(t))\frac{dx}{dt} + q(x(t), y(t))\frac{dy}{dt} \right) dt$$

can be written, using the dot product, as

$$\int_a^b V(\gamma(t)) \cdot \gamma'(t)\,dt = \int_a^b \left(V(\gamma(t)) \cdot \frac{\gamma'(t)}{\|\gamma'(t)\|} \right) \|\gamma'(t)\|\,dt,$$

which is the integral of the projection of the vector field along the tangent to the curve. (For this last formula, we must assume the tangent vector is not zero.) If the vector field represents a force, it is the work done by the force along the curve.

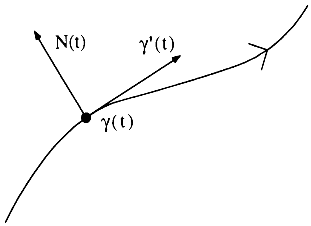
The translation of the equation $\omega = df$, or $p = \partial f/\partial x$ and $q = \partial f/\partial y$, into vector field language says that the corresponding vector field is the *gradient* of f:

$$\text{grad}(f) = \frac{\partial f}{\partial x}\mathbf{i} + \frac{\partial f}{\partial y}\mathbf{j}.$$

The function f (or sometimes $-f$) is called a *potential* function.

We'll finish this chapter with a quick sketch of the interpretation of these ideas for case where the vector field gives the velocity of a fluid flowing in an open set in the plane. So $V(x, y)$ is the velocity vector at (x, y). We assume the flow is in a steady state, which means that, as written, this velocity vector depends only on the point (and not on time). Let p and q be the components of V, as above.

The integral $\int_\gamma p\,dx + q\,dy$ represents the *circulation of the fluid along the path* γ, per unit of time. This can be seen from the interpretation as the integral of the projection of the velocity vector in the tangent direction along the curve. There is another important integral, $\int_\gamma q\,dx - p\,dy$, that represents the *flux of the fluid across the path* γ (per unit of time). To see this, let $N(t) = (-dy/dt, dx/dt)$ be the normal vector, which is perpendicular to and 90° counterclockwise from the tangent vector.



Then

$$\begin{aligned}\int_\gamma q\,dx - p\,dy &= \int_a^b \left(q(x(t), y(t))\frac{dx}{dt} - p(x(t), y(t))\frac{dy}{dt}\right) dt \\ &= \int_a^b V(\gamma(t)) \cdot N(t)\,dt = \int_a^b \left(V(\gamma(t)) \cdot \frac{N(t)}{\|N(t)\|}\right) \|\gamma'(t)\|\,dt.\end{aligned}$$

This is the integral of the projection of the velocity vector in the normal direction, which measures the flow across γ, from right to left in the direction along γ, per unit of time.

The flow is called *irrotational* if the circulation around all small closed loops is zero. By Problem 2.15, this is true precisely when $d\omega = 0$, i.e., $\partial q/\partial x - \partial p/\partial y = 0$. The function $\partial q/\partial x - \partial p/\partial y$ is called the *curl* of the vector field.

The flow is called *incompressible* if the net flow across small closed loops is zero. Applying Problem 2.15 to the 1-form $q\,dx - p\,dy$, this is equivalent to the condition that $\partial p/\partial x + \partial q/\partial y = 0$. The function $\partial p/\partial x + \partial q/\partial y$ is called the *divergence* of the vector field.

Exercise 2.20. If R is a rectangle (or disk) in U, show that the integral of the curl (identified with the 2-form $(\partial q/\partial x - \partial p/\partial y)\,dx\,dy$) over R is the circulation around ∂R, and the integral of the divergence is the flux across ∂R.

If the open set U is a simple one, such as an open rectangle or disk,

and if the flow is irrotational, then we know that there is a potential function f, i.e., $V = \text{grad}(f)$. If the fluid is also incompressible, the equation we just found says that

$$\frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} = 0,$$

which is the condition for f to be a *harmonic* function.

Although we have no topological excuses for them, here are a handful of typical applications of and variations on these notions.

Exercise 2.21. (a) Show that the following functions are harmonic: (i) $a + bx + cy$ for any a, b, c; (ii) $x^2 - y^2$; (iii) $\log(r)$, $r = \sqrt{x^2 + y^2}$ on the complement of the origin. (b) Find all harmonic polynomial functions of x and y of degree at most three. (c) Find all harmonic functions that have the form $h(r)$.

Problem 2.22. (a) State and prove an analogue of Green's theorem when R is the region between two rectangles, one contained in the other. (b) State and prove an analogue of Green's theorem when R is a disk, or the region between two concentric disks.

Problem 2.23. If R is a region, with boundary ∂R, for which Green's theorem is known, prove the following two formulas of Green, for functions f and g on a region containing the closure of R:

$$\text{(i)} \int_{\partial R} f \cdot \left(-\frac{\partial g}{\partial y} dx + \frac{\partial g}{\partial x} dy \right) = \iint_R \left(f \cdot \left(\frac{\partial^2 g}{\partial x^2} + \frac{\partial^2 g}{\partial y^2} \right) + \text{grad}(f) \cdot \text{grad}(g) \right) dx\,dy;$$

$$\text{(ii)} \int_{\partial R} f \cdot \left(-\frac{\partial g}{\partial y} dx + \frac{\partial g}{\partial x} dy \right) - g \cdot \left(-\frac{\partial f}{\partial y} dx + \frac{\partial f}{\partial x} dy \right) = \iint_R \left(f \cdot \left(\frac{\partial^2 g}{\partial x^2} + \frac{\partial^2 g}{\partial y^2} \right) - g \cdot \left(\frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} \right) \right) dx\,dy.$$

Problem 2.24. (a) If f is harmonic in R and vanishes on ∂R, show that f must be identically zero. (b) If two harmonic functions on R have the same restriction to ∂R, show that they must agree everywhere on R.

Problem 2.25. If f is harmonic on a disk, show that the value of f at the center of the disk is the average value of f on the boundary of the disk.

Exercise 2.26. If $V(x,y) = -(1/2\pi)(x,y)/\|(x,y)\|^2$ and γ is a closed path not passing through the origin, show that the flux across γ is the winding number $W(\gamma, 0)$, i.e., $W(\gamma,0) = \int_a^b V(\gamma(t)) \cdot N(t)\, dt$.

These physical notions suggest a way to generalize the notion of a winding number to higher dimensions. If f is a $\mathscr{C}^\infty$ map from a rectangle or disk R to $\mathbb{R}^3 \setminus (0)$, let $N_f(P)$ be the cross product of the columns of the Jacobian matrix of f at P, and let

$$V(x,y,z) = (1/4\pi)(x,y,z)/\|(x,y,z)\|^3.$$

It is a good project to develop a notion of an "engulfing number," setting $W(f,0) = \iint_R V(f(P)) \cdot N_f(P)$.

PART II

WINDING NUMBERS

The notion of winding numbers is generalized to arbitrary continuous paths, and the facts we proved in the smooth case using calculus are proved here by purely topological arguments. We also look at what happens when the point being wound around is varied. Finally, we use the idea of winding numbers to define the degree of a mapping from one circle to another, and to define the local degree of a mapping from one open plane set to another.

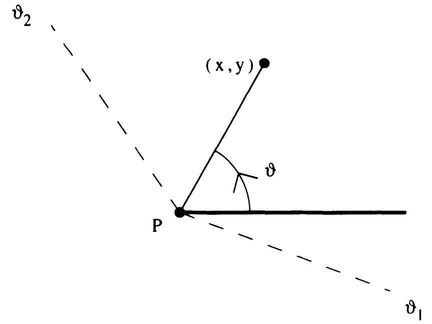
In Chapter 4 there are several applications of winding numbers, some written out, and many left as exercises and problems. Many of these results generalize to higher dimensions; the names attached to some of them refer to these generalizations.

CHAPTER 3

The Winding Number

3a. Definition of the Winding Number

For any point P in the plane, and any sector with vertex at P, we can define a continuous (even $\mathscr{C}^\infty$) angle function on the sector, although the choice is unique only up to adding integral multiples of 2π. Here angles are measured with reference to P, counterclockwise from the horizontal line to the east of P:



If $p = p_P$ is the corresponding polar coordinate map:

$$p(r, \vartheta) \;=\; P + (r\cos(\vartheta), r\sin(\vartheta)),$$

a sector is the image of a strip $\{(r, \vartheta): r > 0 \text{ and } \vartheta_1 < \vartheta < \vartheta_2\}$, where ϑ_1 and ϑ_2 are any real numbers with $0 < \vartheta_2 - \vartheta_1 \leq 2\pi$. The fact that this ϑ is a continuous function on the sector can be seen directly,

using formulas like the arc tangent, or topologically as follows: p is one-to-one and continuous from the strip onto the sector, and it maps small open rectangles $\{(r,\vartheta)\colon a<r<b, \alpha<\vartheta<\beta\}$ onto small open sectors (bounded by two straight lines and arcs of two circles), so p is open as well as continuous; it is therefore a homeomorphism (in fact a diffeomorphism). This means that r and ϑ are continuous functions of x and y.

For any continuous path $\gamma\colon [a,b]\to\mathbb{R}^2\setminus\{P\}$, we define its *winding number* $W(\gamma,P)$ as follows:

Step 1. Subdivide the interval into $a=t_0\le t_1\le \ldots \le t_n=b$, so that each subinterval $[t_{i-1},t_i]$ is mapped into some sector with vertex at P. Such a subdivision exists by the Lebesgue lemma, since each point in the image of γ is contained in some such sector.

Step 2. Choose such a sector U_i containing $\gamma([t_{i-1},t_i])$ and a corresponding angle function ϑ_i on U_i, for $1\le i\le n$. Let $P_i=\gamma(t_i), 0\le i\le n$. Define

$$\begin{aligned} W(\gamma,P) &= \frac{1}{2\pi}[(\vartheta_1(P_1)-\vartheta_1(P_0))+(\vartheta_2(P_2)-\vartheta_2(P_1)) \\ &\qquad + \ldots + (\vartheta_n(P_n)-\vartheta_n(P_{n-1}))] \\ &= \frac{1}{2\pi}\sum_{i=1}^{n}(\vartheta_i(P_i)-\vartheta_i(P_{i-1})). \end{aligned}$$

Each term represents the net change in angle along that part of the path.

Proposition 3.1. (*a*) *The definition of a winding number is independent of the choices made in Steps* 1 *and* 2. (*b*) *If γ is a closed path, i.e., $\gamma(a)=\gamma(b)$, then $W(\gamma,P)$ is an integer.*

Proof. (a) To see that it is independent of the choices of the sectors U_i and the angle functions ϑ_i, suppose U_i' and ϑ_i' were other choices. Then ϑ_i and ϑ_i' would differ by a constant (in fact, a multiple of 2π) on the component of the intersection of U_i and U_i' that contains $\gamma([t_{i-1},t_i])$. So the difference in the values of ϑ_i at the two points P_{i-1} and P_i is the same as the difference in values of ϑ_i' at these two points, and adding over i shows that the winding number doesn't change.

So it is enough to show that the definition is independent of the

choice of subdivision. Suppose we add one point to a given subdivision that satisfies the condition in Step 1, say by inserting a point t^* between some t_{i-1} and t_i. If U_i and ϑ_i are chosen for $[t_{i-1}, t_i]$, these same U_i and ϑ_i can also be chosen for the two subintervals $[t_{i-1}, t^*]$ and $[t^*, t_i]$. And if $P^* = \gamma(t^*)$, then

$$(\vartheta_i(P_i) - \vartheta_i(P^*)) + (\vartheta_i(P^*) - \vartheta_i(P_{i-1})) \;=\; \vartheta_i(P_i) - \vartheta_i(P_{i-1}),$$

so again the sum is unchanged. It follows that if we insert any finite number of points into a given subdivision, the definition of the winding number will not change. But then, if two subdivisions both satisfy the condition in Step 1, the common refinement of both of them, obtained by including all division points for each, will define the same winding number as each of them.

(b) In general, the claim is that, even when γ is not closed, if ϑ_a is an angle for the initial point $\gamma(a)$, then $\vartheta_a + 2\pi \cdot W(\gamma, P)$ is an angle for the endpoint $\gamma(b)$. When the path is closed, this implies that $W(\gamma, P)$ is an integer. This claim is evident for the restriction of γ to each subinterval $[t_{i-1}, t_i]$ on which there is a continuous angle function, and the general case follows by adding up the results (or inducting on the number of subintervals). □

Exercise 3.2. Show that if γ is smooth, this definition agrees with that in Chapter 2.

Exercise 3.3. Show that if $\gamma\colon [a, b] \to U$ is a closed path, and U is an open set in $\mathbb{R}^2 \setminus \{P\}$ on which there is a continuous angle function (for example, a sector with vertex at P), then $W(\gamma, P) = 0$.

Exercise 3.4. Show that winding numbers are invariant by translation, in the following sense. Let $\gamma\colon [a, b] \to \mathbb{R}^2 \setminus \{P\}$, and let v be any vector in the plane. Let $\gamma + v$ be the path defined by $(\gamma + v)(t) = \gamma(t) + v$. Show that

$$W(\gamma + v, P + v) \;=\; W(\gamma, P).$$

Problem 3.5. Show that for any continuous path $\gamma\colon [a, b] \to \mathbb{R}^2 \setminus \{P\}$, there are continuous functions $r\colon [a, b] \to \mathbb{R}^+$ (the positive real numbers) and $\vartheta\colon [a, b] \to \mathbb{R}$, so that

$$\gamma(t) \;=\; P + (r(t)\cos(\vartheta(t)), r(t)\sin(\vartheta(t))), \qquad a \le t \le b.$$

Show that r is uniquely determined, and ϑ is uniquely determined up to adding a constant integral multiple of 2π. Show in fact that $r(t) = \|\gamma(t) - P\|$, and if γ' denotes the restriction of γ to the interval

$[a,t]$ (so $\gamma^t(u)=\gamma(u)$ for $a\le u\le t$), and ϑ_a is an angle for $\gamma(a)$, then one may take

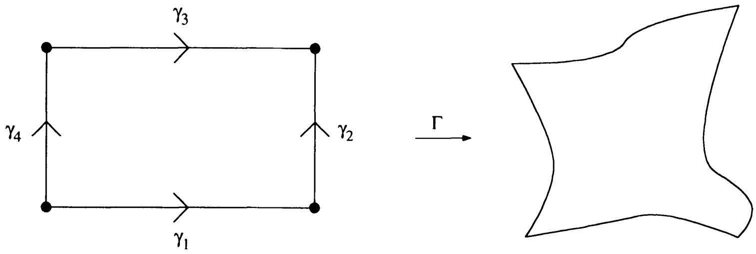
$$\vartheta(t) \;=\; \vartheta_a + 2\pi\cdot W(\gamma^t,P).$$

Equivalently, for any choice of ϑ_a, there is a unique continuous mapping $\tilde\gamma\colon [a,b]\to\{(r,\vartheta)\colon r>0\}$ such that $p_P\circ\tilde\gamma=\gamma$ and $\tilde\gamma(a)=(r(a),\vartheta_a)$, where p_P is the polar coordinate mapping. Such a $\tilde\gamma$ is called a *lifting* of γ with starting point $(r(a),\vartheta_a)$.

3b. Homotopy and Reparametrization

Suppose $R=[a,b]\times[c,d]$ is a closed rectangle, and $\Gamma\colon R\to\mathbb{R}^2$ is a continuous mapping. The restrictions of Γ to the four sides of the rectangle define four paths γ_1, γ_2, γ_3, and γ_4:

$$\begin{aligned}
\gamma_1(t)&=\Gamma(t,c), &\quad \gamma_3(t)&=\Gamma(t,d), &\quad a&\le t\le b;\\
\gamma_2(s)&=\Gamma(b,s), &\quad \gamma_4(s)&=\Gamma(a,s), &\quad c&\le s\le d.
\end{aligned}$$



Theorem 3.6. *For any point P not in $\Gamma(R)$,*

$$W(\gamma_1,P)+W(\gamma_2,P) \;=\; W(\gamma_3,P)+W(\gamma_4,P).$$

Proof. In fact, the second proofs we gave for Proposition 2.16 and 2.17 work equally well in this case. As there, use the Lebesgue lemma to subdivide the rectangle into subrectangles $R_{i,j}$, such that Γ maps $R_{i,j}$ into a sector $U_{i,j}$ (with vertex at P), on which there is a continuous angle function $\vartheta_{i,j}$. Then, by the canceling of inside edges as before,

$$W(\Gamma|_{\partial R},P) \;=\; \sum_{i,j} W(\Gamma|_{\partial R_{i,j}},P),$$

where $W(\Gamma|_{\partial R},P)$ is defined by the equation

$$W(\Gamma|_{\partial R},P) \;=\; W(\gamma_1,P)+W(\gamma_2,P)-W(\gamma_3,P)-W(\gamma_4,P),$$

and $W(\Gamma|_{\partial R_{i,j}}, P)$ is defined similarly as the signed sum of the winding numbers around the small rectangles. But these winding numbers around the small rectangles are all zero, since each $R_{i,j}$ is mapped into a region where there is an angle function $\vartheta_{i,j}$. So

$$W(\gamma_1, P) + W(\gamma_2, P) - W(\gamma_3, P) - W(\gamma_4, P) \;=\; 0. \qquad \square$$

This theorem implies that the winding numbers have the same invariance under homotopies as in the smooth case considered in §2b. Again, there are two kinds of homotopies we want to consider, first for paths with fixed endpoints, and second for closed paths. If $\gamma\colon [a,b] \to U$ and $\delta\colon [a,b] \to U$ are paths with the same initial and final points, a *homotopy from* γ *to* δ *with fixed endpoints* is a continuous mapping $H\colon [a,b] \times [0,1] \to U$ such that

$$H(t,0) = \gamma(t) \quad \text{and} \quad H(t,1) = \delta(t) \qquad \text{for all } a \le t \le b;$$
$$H(a,s) = \gamma(a) = \delta(a) \quad \text{and} \quad H(b,s) = \gamma(b) = \delta(b) \quad \text{for all } 0 \le s \le 1.$$

The paths γ_s defined by $\gamma_s(t) = H(t,s)$ give a continuous family of smooth paths from $\gamma_0 = \gamma$ to $\gamma_1 = \delta$. The paths γ and δ are called *homotopic with fixed endpoints* if there is such a homotopy H.

On the other hand, if γ and δ are closed paths in U, again defined on the same interval $[a,b]$, a *homotopy from* γ *to* δ *through closed paths* is a continuous $H\colon [a,b] \times [0,1] \to U$, such that

$$H(t,0) = \gamma(t) \quad \text{and} \quad H(t,1) = \delta(t) \qquad \text{for all } a \le t \le b;$$
$$H(a,s) = H(b,s) \qquad \text{for all } 0 \le s \le 1.$$

The paths γ and δ are called *homotopic closed paths* if there is such a homotopy H.

Exercise 3.7. Prove that the relation of being homotopic with fixed endpoints, or as closed paths, is an equivalence relation.

Corollary 3.8. *If two paths γ and δ in $\mathbb{R}^2 \setminus \{P\}$ are homotopic, either as paths with the same endpoints, or as closed paths, then*

$$W(\gamma, P) \;=\; W(\delta, P).$$

Proof. This is an immediate consequence of the theorem, applied to the homotopy $H = \Gamma$. In the first case, the winding numbers of the constant paths from the sides of the rectangle are both zero, and in the second case they are the same, so their winding numbers cancel in the result. $\square$

We next consider what happens to the winding number by a change

of parameter, generalizing what we saw in the last chapter in the smooth case.

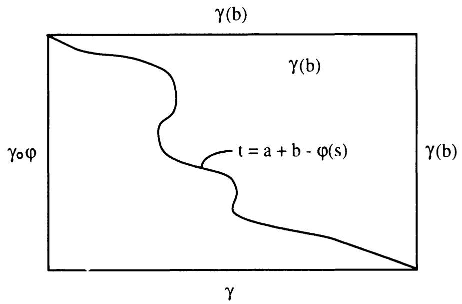
Corollary 3.9. *Let γ: $[a,b] \to \mathbb{R}^2 \setminus \{P\}$ be a continuous path, and φ: $[c,d] \to [a,b]$ a continuous function.*

(a) If $\varphi(c) = a$ and $\varphi(d) = b$, then $W(\gamma \circ \varphi, P) = W(\gamma, P)$.
(b) If $\varphi(c) = b$ and $\varphi(d) = a$, then $W(\gamma \circ \varphi, P) = -W(\gamma, P)$. In particular, if $\gamma^{-1}(t) = \gamma(a+b-t)$, $a \le t \le b$, then

$$W(\gamma^{-1}, P) = -W(\gamma, P).$$

Proof. For (a), define Γ: $[a,b] \times [\mathrm{c}, \mathrm{d}] \to U$ by the formula

$$\Gamma(t,s) = \gamma(\min(t + \varphi(s) - a, b)).$$



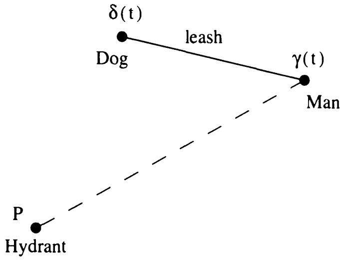
Then Γ is continuous, since the minimum or composite of two continuous functions is continuous. The paths from the sides of Γ are $\gamma_1 = \gamma$, $\gamma_4 = \gamma \circ \varphi$, and γ_2 and γ_3 are constant paths at the point $\gamma(\mathrm{b})$. So Theorem 3.6 applies and gives (a). Similarly for (b), use

$$\Gamma(s,t) = \gamma(\max(t + \varphi(s) - b, a)).$$

In this case $\gamma_1 = \gamma$, $\gamma_2 = \gamma \circ \varphi$, and γ_3 and γ_4 are constant. □

Exercise 3.10. Give a direct proof of Corollary 3.9 from the definition of the winding number, in case the change of coordinates φ is a monotone increasing or monotone decreasing function. (Monotone increasing means that $\varphi(t) < \varphi(s)$ if $t < s$.)

As another application, we have the "dog-on-a-leash" theorem of Poincaré and Bohl. This says if the leash is kept shorter than the distance from the walker to the fire hydrant, then the walker and the dog both wind around the hydrant the same number of times:



Theorem 3.11 (Dog-on-a-Leash). *Suppose* $\gamma\colon [a,b] \to \mathbb{R}^2 \setminus \{P\}$ *and* $\delta\colon [a,b] \to \mathbb{R}^2 \setminus \{P\}$ *are closed paths, and the line segment between* $\gamma(t)$ and $\delta(t)$ *never hits the point P. Then*

$$W(\gamma, P) = W(\delta, P).$$

Proof. Define $H\colon [a,b] \times [0,1] \to \mathbb{R}^2$ by the formula

$$H(t,s) = (1-s)\gamma(t) + s\,\delta(t), \qquad a \le t \le b, \quad 0 \le s \le 1.$$

This is a continuous homotopy from γ to δ through closed paths. The hypotheses imply that Γ maps the rectangle into $\mathbb{R}^2 \setminus \{P\}$. The result therefore follows from Corollary 3.8. □

Corollary 3.12. *If* $\gamma\colon [a,b] \to \mathbb{R}^2$ *and* $\delta\colon [a,b] \to \mathbb{R}^2$ *are closed paths such that* $\|\gamma(t) - \delta(t)\| < \|\gamma(t) - P\|$ *for all t in* $[a,b]$, *then* $W(\gamma, P) = W(\delta, P)$.

Proof. The hypothesis implies that neither path hits P, and that the line segment between $\gamma(t)$ and $\delta(t)$ doesn't hit P. □

Exercise 3.13. Show that if U is an open rectangle (bounded or unbounded), then any two paths from $[a,b]$ to U with the same endpoints are homotopic, and any two closed paths in U are homotopic. Show that the same is true for any convex open set U. Can you prove it when U is just starshaped?

Problem 3.14. Let γ and δ be paths from an interval to $\mathbb{R}^2 \setminus \{P\}$ with the same endpoints. Show that the following are equivalent:

(i) γ and δ are homotopic in $\mathbb{R}^2 \setminus \{P\}$;
(ii) $W(\gamma, P) = W(\delta, P)$; and
(iii) if $\tilde{\gamma}$ and $\tilde{\delta}$ are liftings of γ and δ with the same initial point, as in Problem 3.5, then $\tilde{\gamma}$ and $\tilde{\delta}$ have the same final point.

Problem 3.15. Let γ and δ be closed paths from a closed interval to

$\mathbb{R}^2 \setminus \{P\}$. Show that γ and δ are homotopic through closed paths in $\mathbb{R}^2 \setminus \{P\}$ if and only if $W(\gamma, P) = W(\delta, P)$.

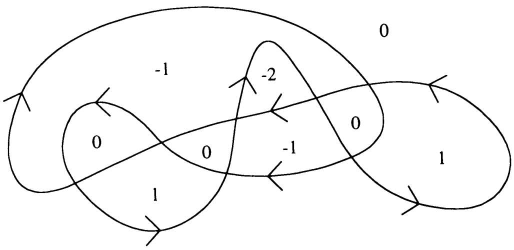
3c. Varying the Point

We want to study what happens to the winding number if a closed path γ: $[a,b] \rightarrow \mathbb{R}^2$ is fixed, and the point P is allowed to vary, but always so that the path does not pass through P. Let us denote the image of the path γ by Supp(γ), and call it the *support* of γ, i.e.,

$$\text{Supp}(\gamma) = \gamma([a,b]).$$

Since the interval is compact, the support is a compact, and hence closed and bounded, subset of the plane. The complement of the support is an open set, which may have many—even infinitely many—connected components, each of which is open. Since the support is bounded, however, there is one connected component of $\mathbb{R}^2 \setminus \text{Supp}(\gamma)$ that contains all points outside some large disk; this component is called the *unbounded* component.

Proposition 3.16. *As a function of P, the function $W(\gamma, P)$ is constant on each connected component of $\mathbb{R}^2 \setminus Supp(\gamma)$. It vanishes on the unbounded component.*



Proof. For the first statement, we must show that $W(\gamma, P)$ is a locally constant function of P in $\mathbb{R}^2 \setminus \text{Supp}(\gamma)$. Given P, choose a disk around P contained in $\mathbb{R}^2 \setminus \text{Supp}(\gamma)$. We must show that $W(\gamma, P') = W(\gamma, P)$ for all P' in the disk. Let $v = P' - P$ be the vector from P to P'. By Exercise 3.4,

$$W(\gamma, P) = W(\gamma + v, P + v) = W(\gamma + v, P').$$

The homotopy $H(t,s) = \gamma(t) + sv$, $a \le t \le b$, $0 \le s, \le 1$, is a homotopy through closed paths from γ to $\gamma + v$ that never hits the point P',

so

$$W(\gamma + v, P') = W(\gamma, P').$$

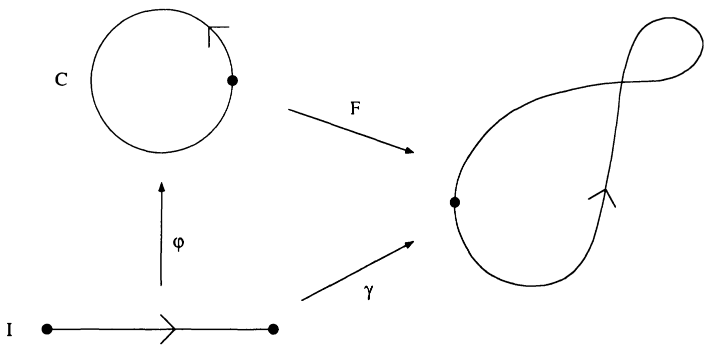
To show that the winding number vanishes on the unbounded component, it suffices to show it vanishes on one such point. For example, we can take P far out on the negative x-axis, so that the support of γ is contained in a half plane to the right of P. But then there is an angle function on this half plane, so the winding number is zero by Exercise 3.3. □

The following exercise gives alternative proofs of the proposition:

Exercise 3.17. (a) Prove directly from the definition that for any path γ, whether closed or not, the function $P \mapsto W(\gamma, P)$ is continuous on the complement of $\text{Supp}(\gamma)$, and approaches zero when $\|P\|$ goes to infinity. (b) Use (a) and the fact that, when γ is closed, $W(\gamma, P)$ takes values in the integers to give another proof of the proposition. (c) For a closed path γ, show directly that $W(\gamma, P)$ is constant on path components of $\mathbb{R}^2 \setminus \text{Supp}(\gamma)$ by applying Theorem 3.6 to the mapping $\Gamma(t, s) = \gamma(t) - \sigma(s)$ and the point $P = 0$, where σ is any path in $\mathbb{R}^2 \setminus \text{Supp}(\gamma)$.

3d. Degrees and Local Degrees

If I is any closed interval, and C is any circle, a continuous closed path γ from I to an open set U is essentially the same thing as a continuous mapping from C into U:



This can be realized explicitly as follows. Let us assume that $I = [0, 1]$,

since this is the most common convention. Suppose C has center (x_0, y_0) and radius r. Let $\varphi: I \to C$ be the function that wraps I around C:

$$\varphi(t) = (x_0, y_0) + (r\cos(2\pi t), r\sin(2\pi t)), \qquad 0 \leq t \leq 1.$$

So φ is a one-to-one mapping of I onto C, except that $\varphi(0) = \varphi(1)$. It follows that if F is a mapping from C into an open set U, then $\gamma = F \circ \varphi$ is a mapping from I to U with $\gamma(0) = \gamma(1)$, and that any such γ can be realized in this way for a unique F.

Lemma 3.18. *The mapping γ is continuous if and only if F is continuous.*

Proof. In fact, φ realizes C as the quotient space of I with its endpoints 0 and 1 identified. Concretely, this means that a subset X of C is open in C if and only if $\varphi^{-1}(X)$ is open in I. This is easy to verify directly, or one can argue that the quotient space is a compact Hausdorff space, and the induced mapping from it to C, being continuous and one-to-one, must be a homeomorphism. If follows that for an open subset V of U, $F^{-1}(V)$ is open if and only if $\gamma^{-1}(V) = \varphi^{-1}(F^{-1}(V))$ is open, which proves the lemma. □

For any continuous $F: C \to \mathbb{R}^2 \setminus \{P\}$, we can define the *winding number of F around P*, denoted $W(F, P)$, to be the winding number $W(\gamma, P)$ of the path $\gamma = F \circ \varphi$.

Exercise 3.19. Identify $\mathbb{R}^2$ with the complex numbers $\mathbb{C}$, and let $f: \mathbb{C} \to \mathbb{C}$ be the mapping that takes a complex number z to its nth power z^n, where n is an integer. Let C be any circle centered at the origin, and let F be the restriction of f to C. Show that $W(F, 0) = n$. If $f(z) = -z$, show that $W(F, 0) = 1$.

Proposition 3.20. *Suppose C is the boundary of the closed disk D, and $F: C \to \mathbb{R}^2 \setminus \{P\}$ extends to a continuous function from D to $\mathbb{R}^2 \setminus \{P\}$. Then $W(F, P) = 0$.*

Proof. If D is the disk of radius r about the point (x_0, y_0) as above, and $\gamma: [0, 1] \to \mathbb{R}^2 \setminus \{P\}$ is the path corresponding to F, and $\widetilde{F}: D \to \mathbb{R}^2 \setminus \{P\}$ is such an extension of F, then

$$H(t, s) = \widetilde{F}\big((x_0, y_0) + s(r\cos(2\pi t), r\sin(2\pi t))\big), \qquad 0 \leq t \leq 1, \quad 0 \leq s \leq 1,$$

gives a homotopy from γ to the constant path at the point $\widetilde{F}((x_0, y_0))$. This homotopy stays inside $\mathbb{R}^2 \setminus \{P\}$, and since the winding number of a constant path is zero, the claim follows from Corollary 3.8. □

Problem 3.21. If F_0 and F_1 are mappings from a circle C to U, corresponding to paths γ_0 and γ_1 from $[0, 1]$ to U, show that γ_0 and γ_1 are homotopic through closed paths if and only if F_0 and F_1 are *homotopic* mappings, i.e., there is a continuous mapping

$$H: C \times [0, 1] \rightarrow U$$

with $H(P \times 0) = F_0(P)$ and $H(P \times 1) = F_1(P)$ for all P in C.

Problem 3.22. Show that the converse of Proposition 3.20 is true: if $W(\gamma, P) = 0$, then γ has a continuous extension to a map from D to $\mathbb{R}^2 \setminus \{P\}$.

Problem 3.23. Let C be a circle centered at the origin, and let $F: C \rightarrow \mathbb{R}^2$ be a continuous mapping such that the vector $F(P)$ is never tangent to the curve C at P, i.e., the dot product $F(P) \cdot P$ is not zero for all P in C. Show that $W(F, 0) = 1$.

The same idea lets us define the *degree* of any continuous mapping F from one circle C to another circle C', which measures how many times the first circle is wound around the second by F. Let P' be the center of the circle C', and define the degree of F by the formula

$$\deg(F) = W(F, P').$$

Exercise 3.24. Show that one could take any point inside C' in place of P'.

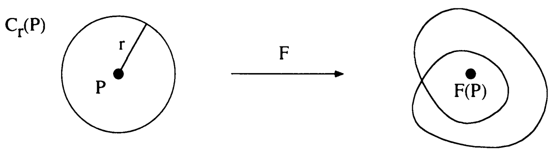
Exercise 3.25. (a) Show that if F is not surjective, then its degree is zero. (b) Give an example of a surjective mapping from C to C' that has degree zero. (c) Show that if F and G are homotopic mappings from C to C', i.e., if there is a continuous mapping H from $C \times [0, 1]$ to C', with $F(Q) = H(Q \times 0)$ and $G(Q) = H(Q \times 1)$ for all Q in C, then F and G have the same degree. (d) Show that if C is the boundary of the disk D, then F extends to a continuous mapping from D to C' if and only if F is homotopic to a constant mapping from C to C', and then the degree of F is zero.

Problem 3.26. (a) Prove that two mappings from C to C' have the same degree if and only if they are homotopic. (b) Deduce that, if C is the boundary of the disk D, and the degree of a mapping is zero, then the mapping extends to a continuous mapping from D to C'. (c) Deduce also that if S^1 is the unit circle centered at the origin, and $F: S^1 \rightarrow S^1$ is a continuous mapping with degree n, then F is homo-

topic to the mapping that takes $(\cos(\vartheta), \sin(\vartheta))$ to $(\cos(n\vartheta), \sin(n\vartheta))$, i.e., in the terminology of complex numbers, the restriction of the nth power mapping $z \mapsto z^n$ to the unit circle.

Problem 3.27. If $F: C \to C'$ and $G: C' \to C''$ are continuous mappings of circles, what can you say about the relation among the degrees of F, G, and the composite $G \circ F$? Can you prove your answer?

These ideas can also be used to define an important notion of a *local degree*. Suppose U and V are open sets in the plane, and $F: U \to V$ is a continuous mapping, and let P be a point in U. Assume that P has some small neighborhood such that $F(Q) \neq F(P)$ for all Q in that neighborhood with $Q \neq P$. Choose a positive number r so that no point within a distance r of P has the same image as P, and let $C_r(P)$ be the circle of radius r about P. Then F restricts to a continuous mapping from $C_r(P)$ to $\mathbb{R}^2 \setminus \{F(P)\}$.



Define the *local degree of F at P*, denoted $\deg_P(F)$, to be the winding number of this mapping of the circle around the point $F(P)$. In other words, $\deg_P(F) = W(\gamma_r, F(P))$, where

$$\gamma_r(t) = F(P + r(\cos(2\pi t), \sin(2\pi t))), \qquad 0 \leq t \leq 1.$$

To know that this is well defined, we need

Lemma 3.28. *This winding number is independent of choice of r.*

Proof. If r' is another, $H(t, s) = F(P + ((1-s)r + sr')(\cos(2\pi t), \sin(2\pi t)))$ gives a homotopy from γ_r to $\gamma_{r'}$. □

Equivalently, the local degree of F at P is the winding number of the mapping from the unit circle S^1 to itself given by

$$Q \mapsto \frac{F(P + rQ) - F(P)}{\|F(P + rQ) - F(P)\|}.$$

Problem 3.29. Show that if $F: \mathbb{R}^2 \to \mathbb{R}^2$ is a linear mapping, given

by a (2×2)-matrix, and the determinant is not zero, then the local degree of F at the origin is $+1$ if this determinant is positive, and -1 if this determinant is negative.

Problem 3.30. Show that if $F: U \to V$ is a $\mathscr{C}^\infty$ mapping, and the Jacobian determinant of F is not zero at P, then the local degree is defined, and is $+1$ or -1, depending on the sign of this determinant.

Problem 3.31. Suppose $F: \mathbb{C} \to \mathbb{C}$ is given by a complex polynomial. Show that the local degree of F at a complex number z is the multiplicity of z as a root of $F(T) - F(z)$. (This multiplicity is the highest power of $T - z$ that divides $F(T) - F(z)$.)

Problem 3.32. If $F: C \to C'$ is a map between circles in the plane, and P is a point in C such that $F(Q) \neq F(P)$ for all Q in a neighborhood of P in C, give a precise definition of a local degree of F at P; this should be $+1$ if F is increasing at P, -1 if F is decreasing, and 0 if F has a local maximum or minimum at P (all expressed in terms of counterclockwise angles). Show that if P' is any point of C' such that $F^{-1}(P')$ is finite, then

$$\deg(F) = \sum_{P \in F^{-1}(P')} \deg_P F,$$

where $\deg_P F$ is the local degree of F at P. This implies in particular that the right side is independent of choice of P'.

CHAPTER 4

Applications of Winding Numbers

4a. The Fundamental Theorem of Algebra

The set $\mathbb{C}$ of complex numbers is identified as usual with the real plane $\mathbb{R}^2$, the number $z = x + iy$ being identified with the point (x, y). We will use the fact that for any complex polynomial

$$g(T) = a_0T^n + a_1T^{n-1} + \ . \ . \ . \ + a_{n-1}T + a_n,$$

with coefficients a_i complex numbers, the mapping $z \mapsto g(z)$ is a continuous mapping from $\mathbb{C}$ to $\mathbb{C}$. This follows from the fact that addition and multiplication of complex numbers are continuous. The goal of this section is to show that, if $n > 0$ and $a_0 \neq 0$, then the polynomial has a root, i.e., $g(z) = 0$ for some z. We may divide by a_0, so we can assume $g(T)$ has leading coefficient $a_0 = 1$. If $g(T)$ has no root, g is a mapping of $\mathbb{C}$ into the complement of the origin.

Restrict g to a circle C_r of radius r centered at the origin. This gives a mapping from C_r to $\mathbb{C} \setminus \{0\}$, which we denote by g_r. Since g_r extends to a continuous mapping of the disk D_r of radius r into $\mathbb{C} \setminus \{0\}$, it follows from Proposition 3.20 that the winding number $W(g_r, 0)$ must be zero. The idea is to compare g_r with the mapping f_r given similarly by the polynomial $f(T) = T^n$. The restriction of this to the circle is $f_r(z) = z^n$, and a simple calculation shows that $W(f_r, 0) = n$ (see Exercise 3.19). We will apply the Dog-on-a-Leash Theorem 3.11 to show that, for r sufficiently large, f_r and g_r must have the same winding number, which will be the desired contradiction. For this, it suffices

to show that for r sufficiently large,

$$|f_r(z) - g_r(z)| < |f_r(z) - 0| \qquad \text{for} \quad z \in C_r.$$

Now $|f_r(z) - 0| = |z^n| = r^n$, and

$$\begin{aligned} |f_r(z) - g_r(z)| &= |a_1 z^{n-1} + \ldots + a_{n-1}z + a_n| \\ &\leq |a_1| r^{n-1} + \ldots + |a_{n-1}| r + |a_n|, \end{aligned}$$

which is less than r^n if r is large, e.g., if $|a_i| < r^i/n$ for all i. This completes the proof of

Proposition 4.1 (Fundamental Theorem of Algebra). *Any complex polynomial of degree greater than zero has a root.*

If z_1 is a root of $g(T)$, then $g(T) = (T - z_1) \cdot h(T)$, where $h(T)$ is a polynomial of degree $n - 1$. By induction, we see that $g(T)$ factors into linear factors: $g(T) = a_0 \cdot \prod_{i=1}^{n}(T - z_i)$.

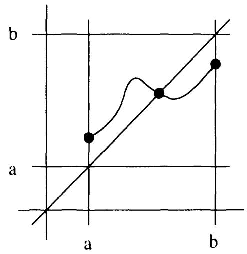
Exercise 4.2. (a) Suppose $f: \mathbb{C} \to \mathbb{C}$ is a continuous function such that for some $R > 0$, $|f(z)| < |z|^n$ for $|z| = R$. Show that $z^n + f(z) = 0$ has a solution z with $|z| < R$. (b) Suppose $g: \mathbb{C} \to \mathbb{C}$ is a continuous function such that $g(z)/z^n$ approaches a nonzero constant as $|z| \to \infty$. Show that g is surjective.

There is another topological proof of the Fundamental Theorem of Algebra that is in some ways even simpler, but requires a little more about the local structure of a mapping given by a polynomial. One shows that g extends to a continuous mapping of the Riemann sphere to itself, taking the point at infinity to itself (which is essentially what the above calculation showed), and which is an open mapping (see Problem 3.31, and §19a for details). The image is compact, so closed, and since both open and closed, it is the whole sphere.

4b. Fixed Points and Retractions

One of the most important applications of topological ideas in other areas of mathematics, and in science in general, is to give criteria to guarantee that a continuous mapping from a space to itself must have a point that is mapped to itself.

We start with the case of a closed interval $[a, b]$. We claim that any continuous function $f: [a, b] \to [a, b]$ must have a fixed point. This can be "seen" by looking at the graph of the mapping:



To prove it rigorously, consider the function $g(x) = f(x) - x$. This is a continuous function on the interval $[a, b]$, with $g(a) \geq 0$ and $g(b) \leq 0$. Since the image $g([a, b])$ of the interval by g must be connected, it must contain the interval $[g(b), g(a)]$, so it must contain 0. This means that $g(x) = 0$ for some x, which says that $f(x) = x$.

This is closely related to another property of an interval: that there is no continuous mapping from an interval $[a, b]$ onto its endpoints $[a, b]$ that maps a to a and b to b. This fact is obvious since the image of a connected set must be connected. In general if Y is a subspace of a topological space X, a *retraction* from X to Y is a continuous mapping $r\colon X \to Y$ such that $r(P) = P$ for all P in Y. In this case Y is called a *retract* of X. So there is no continuous retraction of an interval onto its boundary. This proves again that any continuous function f from the interval $[-1, 1]$ to itself must have fixed points, for otherwise the mapping $x \mapsto (x - f(x))/|x - f(x)|$ would be a continuous retraction of $[-1, 1]$ onto $\{-1, 1\}$.

Next we turn to the case of a closed disk D. We first show there is no retraction onto its boundary circle $C = \partial D$, and then use this as above to show that any map from the disk to itself must have a fixed point.

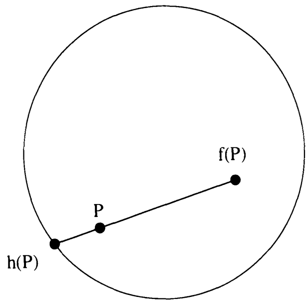
Proposition 4.3. *There is no retraction from a closed disk onto its boundary circle.*

Proof. If C is the boundary of the disk D, the identity mapping from C to itself has degree 1. A retraction would give an extension of the identity mapping to a mapping from D to C, which would imply by Proposition 3.20 that the degree is 0. □

Proposition 4.4 (Brouwer). *Any continuous mapping from a closed disk to itself must have a fixed point.*

Proof. We may assume the disk D is the unit disk centered at the origin, with boundary C the unit circle (see Exercise 4.7 below). Sup-

pose $f: D \to D$ is a continuous mapping with no fixed point. The idea is to define a mapping $h: D \to C$ that takes a point P to the point in C hit by the ray from $f(P)$ to P:



This mapping h will be the identity on C, so will be a retraction of D onto C. To finish the proof, we must verify that h is continuous. To do this, we find an explicit formula for h. We know that

$$h(P) = P + t \cdot U \qquad \text{with} \quad U = \frac{P - f(P)}{\|P - f(P)\|},$$

and t is the positive number determined by the property that $\|h(P)\| = 1$. A little calculation shows that

$$t = -P \cdot U + \sqrt{1 - P \cdot P + (P \cdot U)^2}$$

is such a positive number. With this t, the above formula for h is then continuous, and is easily checked to be the identity on C. □

The following exercise gives a proof with less calculation:

Exercise 4.5. If $f: D^2 \to D^2$ has no fixed point, define $g: D^2 \to \mathbb{R}^2 \setminus \{0\}$ by setting $g(P) = P - f(P)$. Show that $g(P) \cdot P > 0$ for all P in S^1, so the restriction of g to S^1 is homotopic to the identity mapping of S^1 (by the homotopy $H(P \times s) = (1-s)P + s\,g(P)$). Apply Proposition 3.20.

Exercise 4.6. Deduce Proposition 4.3 from Proposition 4.4.

One says that a topological space X *has the fixed point property* if every continuous mapping $f: X \to X$ has some fixed point, i.e., there is a point P in X with $f(P) = P$. So any closed interval and any closed disk have the fixed point property.

Exercise 4.7. Show that a space that is homeomorphic to a space

with the fixed point property has the fixed point property. Show that if Y is a retract of a space X that has the fixed point property, then Y has the fixed point property.

Exercise 4.8. Which of the following spaces have the fixed point property? (i) a closed rectangle; (ii) the plane; (iii) an open interval; (iv) an open disk; (v) a circle; (vi) a sphere S^2; (vii) a torus $S^1 \times S^1$; and (viii) a solid torus $S^1 \times D^2$.

Exercise 4.9. Let D be a disk with boundary circle C, and let $f: D \to \mathbb{R}^2$ be a continuous mapping. Suppose P is a point in $\mathbb{R}^2$ that is not in the image C, and the winding number of the restriction $f|_C$ of f to C around P is not zero. Show that there is some point Q in D such that $f(Q) = P$.

Exercise 4.10. Suppose D and D' are disks with boundary circles C and C'. Suppose $f: D \to \mathbb{R}^2$ is a continuous mapping that maps C into C', such that the degree of this map from C to C' is not zero. Show that $f(D)$ must contain D'.

Exercise 4.11. Show that if $f: D^2 \to \mathbb{R}^2 \setminus \{0\}$ is a continuous mapping, there is some P in $S^1 = \partial D^2$ and some $\lambda > 0$ such that $f(P) = \lambda \cdot P$, and there is some Q in S^1 and some $\mu < 0$ such that $f(Q) = \mu \cdot Q$.

Exercise 4.12. Suppose F is a continuous mapping from the positive octant $\{(x, y, z): x \geq 0, y \geq 0, z \geq 0\}$ to itself. Show that there is a unit vector P in this octant, and a nonnegative number λ, such that $F(P) = \lambda \cdot P$.

Exercise 4.13. If all the entries of a (2×2)-matrix are nonnegative, show by direct calculation that at least one of its eigenvalues must be nonnegative. Prove the same for a (3×3)-matrix A.

Exercise 4.14. If $f: D^2 \to \mathbb{R}^2$ is a continuous mapping, show that either: (i) there is either some point $Q \in D^2$ such that $f(Q) = Q$; or (ii) there is some P_1 in S^1 and some $\lambda_1 > 1$ such that $f(P_1) = \lambda_1 \cdot P_1$, and there is some point P_2 in S^1 and some $\lambda_2 < 1$ such that $f(P_2) = \lambda_2 \cdot P_2$.

Exercise 4.15. If $f: C \to C$ is a continuous mapping with no fixed point, show that degree of f must be 1. In particular, if f has no fixed point, show that f must be surjective.

Exercise 4.16. If $f: C \to \mathbb{R}^2$ is continuous and $W(f, P) \neq 0$, show that every ray emanating from P meets $f(C)$.

Exercise 4.17. If $f: D^2 \to \mathbb{R}^2$ is continuous and $f(P) \cdot P \neq 0$ for all P in S^1, show that there is some Q in D^2 with $f(Q) = 0$.

Problem 4.18. Let $D^\infty = \{(a_0, a_1, a_2, \ldots): \Sigma_{n=0}^\infty a_n^2 \leq 1\}$, the unit ball in the metric space of infinite sequences such that $\Sigma_{n=0}^\infty a_n^2 < \infty$. (a) Find a continuous mapping $f: D^\infty \to D^\infty$ that has no fixed point. (b) Find a continuous retraction of D^∞ onto $S^\infty = \{(a_0, a_1, \ldots): \Sigma_{n=0}^\infty a_n^2 = 1\}$.

4c. Antipodes

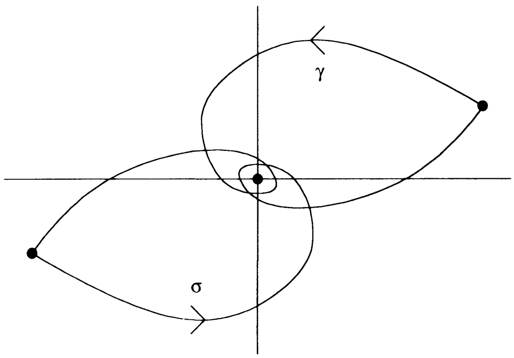
The *antipode* of a point in a circle C or sphere S is the opposite point, i.e., the point hit by the ray from the point through the center. We denote the antipode of P by P^*. With the center at the origin, $P^* = -P$. The *antipodal map* is the mapping that takes each point to its antipode.

Exercise 4.19. Show that the degree of the antipodal map from a circle to itself is 1.

Lemma 4.20 (Borsuk). *If C and C' are circles, and $f: C \to C'$ is a continuous map such that $f(P^*) = f(P)^*$ for all P, then the degree of f is odd.*

Proof. There is no loss of generality in assuming $C = C' = S^1$. Let $\gamma(t) = f(\cos(t), \sin(t)), 0 \leq t \leq \pi$. Since $\gamma(\pi) = -\gamma(0)$, the change in angle along γ must be $2\pi n + \pi$ for some integer n. So $W(\gamma, 0) = n + {}^1\!/_2$. Now let

$$\sigma(t) = f(\cos(t + \pi), \sin(t + \pi)) = -\gamma(t), \qquad 0 \leq t \leq \pi.$$



It follows easily from the definition of the winding number that $W(\sigma, 0) = W(\gamma, 0)$, and so $\deg(f) = W(\gamma, 0) + W(\sigma, 0) = 2n + 1$. □

Lemma 4.21. *There is no continuous mapping f from a sphere S to a circle C such that $f(P^*) = f(P)^*$ for all P in S.*

Proof. Again we can take $S = S^2$. Consider the mapping $g\colon D^2 \to C$ given by the formula

$$g(x, y) = f((x, y, \sqrt{1 - x^2 - y^2})).$$

This is continuous, since the projection from the upper hemisphere of the sphere to the disk (being one-to-one and continuous) is a homeomorphism, and g is the composite of f with the inverse. Now $g(P^*) = f(P^*) = f(P)^* = g(P)^*$ for P in S^1, so by Lemma 4.20 the degree of the restriction of g to S^1 must be odd. But since this map extends over the disk, its degree must be zero, a contradiction. □

Proposition 4.22 (Borsuk–Ulam). *For any continuous mapping $f\colon S \to \mathbb{R}^2$ from a sphere S to the plane, there is a point P in S such that $f(P) = f(P^*)$.*

So there are always two antipodal points on the earth with the same temperature and humidity—or any other two real values, provided they vary continuously over the earth.

Proof. We may take $S = S^2$. If there is no such P, consider the function $g\colon S^2 \to S^1$ given by

$$g(P) = \frac{f(P) - f(-P)}{\|f(P) - f(-P)\|}.$$

Then $g(-P) = -g(P)$, contradicting Lemma 4.21. □

A fact which was unquestionably obvious until people started looking for a proof is the fact that open sets of different dimensions cannot be homeomorphic. (The fact that they cannot be *diffeomorphic* can be reduced, using Jacobian matrices, to the fact that vector spaces of different dimensions cannot be isomorphic.) The fact that there are continuous maps from intervals onto squares makes the assertion less obvious than might have been thought. The first case is easy: an open interval cannot be homeomorphic to an open set in the plane or any $\mathbb{R}^n$, for the reason that removing a point disconnects an interval, but does not disconnect an open set in $\mathbb{R}^n$, $n \geq 2$. The next case is less obvious:

Corollary 4.23 (Invariance of Dimension). *An open set in $\mathbb{R}^2$ cannot be homeomorphic to an open set in $\mathbb{R}^n$ for $n \geq 3$.*

Proof. In fact, no subset of the plane can be homeomorphic to a set which contains a solid ball in $\mathbb{R}^n$, $n \geq 3$, for a homeomorphism from a ball D^n (of some small radius) in the plane would embed a two-sphere $S^2 \subset D^3 \subset D^n$ in the plane, contradicting the proposition. □

As you must expect, these results are also true in higher dimensions, but more machinery is needed to extend the proofs. We'll come back to this in the last part of the book.

Exercise 4.24. Show that if $f: C \to C'$ is a map between circles such that $f(P^*) = f(P)$ for all P, then the degree of f is even.

Exercise 4.25. If $f: C \to \mathbb{R}^2 \setminus \{Q\}$ is a continuous mapping such that Q lies on the line segment between $f(P)$ and $f(P^*)$ for all P in C, show that the winding number of f around Q is odd.

Exercise 4.26. Suppose D is a disk with boundary C, and $f: D \to \mathbb{R}^2$ is a continuous mapping such that $f(P^*) = -f(P)$ for all P in C. Show that there is some point Q in D with $f(Q) = 0$.

Exercise 4.27. If f and g are continuous real-valued functions on a sphere S such that $f(P^*) = -f(P)$ and $g(P^*) = -g(P)$ for all P, show that f and g must have a common zero on the sphere.

Exercise 4.28. State and prove the analogue of the Borsuk–Ulam theorem for mappings from a circle to $\mathbb{R}$.

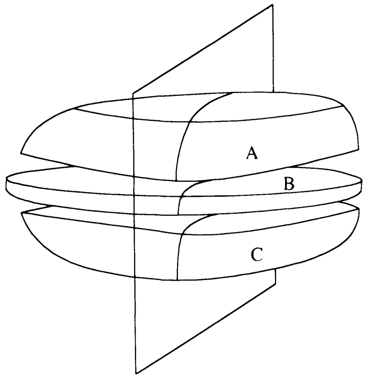
Exercise 4.29. If $f: C \to C'$ is a continuous mapping between circles such that $f(P^*) \neq f(P)$ for all P, show that $\deg(f) \neq 0$, so f must be surjective.

Problem 4.30. Let $f: S \to S'$ be a continuous mapping between spheres. Show that if $f(P) \neq f(P^*)$ for all P, then f must be surjective.

Exercise 4.31. Let $f: C \to C$ be continuous. If $\deg(f) \neq 1$, show that there is a P in C with $f(P) = P$ and there is a Q in C with $f(Q) = Q^*$.

4d. Sandwiches

A pleasant application of the Borsuk–Ulam theorem is the so-called ham sandwich problem: to cut both slices of bread and the ham between in two equal parts with one slice of a knife.



Suppose we have three bounded objects A, B, C in space. The problem is to show that there is one plane that cuts each of them in half (by volume). Here is what we need to know about the volume of each object $X = A$, B, or C:

(i) For any fixed line l, there is a unique point in l such that the plane perpendicular to l through the point cuts X in half. Call this point $P_{l,X}$.
(ii) The point in (i) varies continuously as the line varies continuously. More precisely, take a big sphere S with X inside, and consider for each $Q \in S$ the line $l(Q)$ going from Q to its antipodal point. Then the map from Q to $P_{l(Q),X}$ is continuous.

These properties are intuitively evident, and follow from elementary properties of volume. For example, the first follows from the fact that the volume on one side increases continuously as the point moves along the line. We will take them as axioms, or consider only objects for which we know them. (For a detailed discussion, see Chinn and Steenrod (1966).)

Given a body X inside a sphere S as above, define a continuous real-valued function $f_X: S \to \mathbb{R}$ by defining $f_X(Q)$ to be the distance from Q to the point $P_{l(Q),X}$. Note that $f_X(Q^*) = d - f_X(Q)$, where d is the diameter of the sphere, and Q^* is the antipodal point to Q.

Now for three bodies, take S containing all three, and consider the mapping $g\colon S \to \mathbb{R}^2$ given by

$$g(Q) = (f_A(Q) - f_C(Q), f_B(Q) - f_C(Q)).$$

Then $g(Q^*) = -g(Q)$ for all Q, so by Proposition 4.22 some Q must be mapped to the origin, which means that $P_{l(Q),A} = P_{l(Q),B} = P_{l(Q),C}$, as required. This proves

Proposition 4.32 (Stone–Tukey). *Given three bounded measurable objects A, B, and C in space, there is a plane that divides each in half by volume.*

The same idea is used in the following proposition, which looks quite different.

Proposition 4.33 (Lusternik–Schnirelman–Borsuk). *It is impossible to cover a sphere with three closed sets, none of which contains a pair of antipodal points.*

Proof. Suppose the sphere S is covered by three such closed sets K_1, K_2, and K_3. For each i, define a real-valued continuous function f_i on S, whose value at P is the minimum distance from P to K_i (see the following exercise). Consider the mapping $g\colon S \to \mathbb{R}^2$ given by

$$g(P) = (f_1(P) - f_3(P), f_2(P) - f_3(P)).$$

By Proposition 4.22, there is a point P with $g(P^*) = g(P)$. For such P, $f_i(P) - f_j(P) = f_i(P^*) - f_j(P^*)$ for all i and j. But P must be in one of the sets K_i, and P^* in another K_j. Since $P \notin K_j$ and $P^* \notin K_i$,

$$0 > -f_j(P) = f_i(P) - f_j(P) = f_i(P^*) - f_j(P^*) = f_i(P^*) > 0,$$

a contradiction. □

Exercise 4.34. Show that for any compact set K in space, the function ρ on $\mathbb{R}^3$ that maps a point P to its distance from K is continuous. Note by compactness that if $\rho(P) = r$, then there is a point Q in K of distance r from P.

Exercise 4.35. Show that the unit ball D^3 is not the union of three closed sets, each with diameter less than 2.

Exercise 4.36. Show that the "three" in Proposition 4.33 cannot be replaced by "four."

Exercise 4.37. State and prove the analogue of Proposition 4.33 for a circle.

Exercise 4.38. Show how Proposition 4.33 implies Lemma 4.21, and hence Proposition 4.22.

Exercise 4.39. For a subset X of a sphere, let $X^* = \{P^*: P \in X\}$. Suppose A, B, and C are disjoint closed subsets in a sphere, none containing a pair of antipodal points. Show that the six sets A, B, C, A^*, B^*, and C^* cannot cover the sphere.

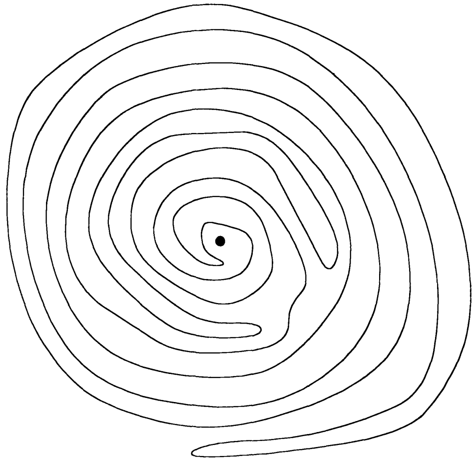
Problem 4.40. True/False. If a sphere is a union of two closed sets A and B, then either A or B must contain a closed connected set X such that $X^* = X$.

It is an excellent project to try to generalize the results of this chapter, including the exercises, to higher dimensions, assuming, for example, that there is no continuous retraction from D^n onto S^{n-1}, and that there is no continuous mapping f from the sphere S^n to the sphere S^{n-1} such that $f(P^*) = f(P)^*$ for all P. This will be carried out in Chapter 23.

PART III
COHOMOLOGY AND HOMOLOGY, I

We have seen that the topology of an open set U in the plane is related to the question of whether all closed 1-forms on U are exact. This is formalized by introducing the vector space H^1U of closed forms modulo exact forms. What we learned in the first few chapters amounts to some calculations of these first De Rham cohomology groups. There is also a 0th De Rham group H^0U, which measures how many connected components U has.

A central theme will be how the topology of a union $U \cup V$ of two open sets compares with the topology of U and V and the intersection $U \cap V$. We construct a linear map from $H^0(U \cap V)$ to $H^1(U \cup V)$, and describe the kernel and cokernel of this map. We use these groups and this map to prove the famous Jordan curve theorem, which says that any subset of the plane homeomorphic to a circle separates the plane into exactly two connected pieces, an "inside" and an "outside."



This is one of those facts that seem intuitively evident, although a complicated enough maze may raise a few doubts. In fact, it went unquestioned until mathematicians realized that continuous mappings can be pretty horrible—for example, that there can be a continuous mapping from an interval onto a square. When X is a closed polygon, it is not too hard to give an elementary proof, and you might enjoy seeing if you can.

A basic question, which motivates Chapter 6 and the continuation in Chapter 9, stems from three different ways one can compare two closed paths γ and δ in an open set U in the plane, supposing for simplicity that they are differentiable:

(1) Are the winding numbers of γ and δ around all points not in U the same?
(2) Are all integrals of closed 1-forms ω on U along γ and δ the same?
(3) Are the paths homotopic, or related by deformations of some kind?

In studying questions like this, we will find it useful to generalize and formalize some of the ideas of previous chapters. In Chapter 6, paths and segmented paths are generalized to the notion of 1-chains, which are arbitrary sums and differences of (nonconstant) paths. We define what it means for two such chains to be homologous: the difference should be a linear combination of boundaries of maps of a rectangle into the region. For an open set in the plane, we show that this notion is equivalent to saying that the two chains have the same winding number around all points outside the open set. The main technique is to approximate general paths by rectangular paths along

sides of a grid.[3] We introduce the first homology group H_1U, which is the group of closed 1-chains up to homology, and a 0th group H_0U, which is a free abelian group on the connected components of U. In Part V we will develop general tools for computing these homology and cohomology groups.

[3] The use of grids in these chapters follows ideas of L.E.J. Brouwer, E. Artin, and L. Ahlfors, see Ahlfors (1979).

CHAPTER 5

De Rham Cohomology and the Jordan Curve Theorem

5a. Definitions of the De Rham Groups

Define, for an open set U in the plane: the *zeroth De Rham group*,

$$H^0U = \{\text{locally constant functions on } U\}.$$

This is a vector space, by the ordinary addition of functions, and multiplication of functions by real scalars. As we have seen, to give a locally constant function on U is the same as giving a constant on each connected component of U. If U has n connected components, then H^0U is an n-dimensional vector space; if $U_1, \ldots, U_n$ are the connected components of U, and e_i is the function that is 1 on U_i and 0 on U_j for all $j \neq i$, then $e_i, \ldots, e_n$ is a basis for H^0U. Similarly, if U has infinitely many components, H^0U is an infinite-dimensional vector space.

The closed 1-forms on U also form a vector space, since the sum of closed forms, and a constant times a closed 1-form, are also closed. The exact 1-forms on U are a subspace of this vector space, since $d(f_1 + f_2) = df_1 + df_2$ and $d(cf) = cdf$. Whenever we have a subspace W of a vector space V, we can form the quotient space V/W of equivalence classes: two vectors in V are equivalent if their difference is in W (see Appendix C). Define the *first De Rham cohomology group of* U, denoted H^1U, by

$$H^1U = \{\text{closed 1-forms on } U\}/\{\text{exact 1-forms on } U\}.$$

If ω is a 1-form on U, we may write $[\omega]$ for the equivalence class in H^1U containing the 1-form ω.

Later we will have general methods for calculating H^1U. In this section we use the ideas of Part I to compute a few simple examples. For any point $P = (x_0, y_0)$, let ω_P denote the 1-form

$$\omega_P = \frac{1}{2\pi}\omega_{P,\vartheta} = \frac{1}{2\pi}\frac{-(y-y_0)\,dx + (x-x_0)\,dy}{(x-x_0)^2 + (y-y_0)^2}.$$

We have seen that ω_P is a closed 1-form on any open set not containing P.

Proposition 5.1. *(a) If U is an open rectangle, then $H^1U = 0$. (b) If $U = \mathbb{R}^2 \setminus \{P\}$, then $[\omega_P]$ is a basis for H^1U. (c) If $U = \mathbb{R}^2 \setminus \{P, Q\}$, with $P \neq Q$, then $[\omega_P]$ and $[\omega_Q]$ form a basis for H^1U.*

Proof. Assertion (a) is a translation of Proposition 1.12. To prove (b), fix a positive number r, and let $\gamma_{P,r}$ denote the counterclockwise circle of radius r about P:

$$\gamma_{P,r}(t) = P + r(\cos(2\pi t), \sin(2\pi t)), \qquad 0 \leq t \leq 1.$$

Note first that $[\omega_P]$ is not 0 in H^1U, for if ω_P were exact, its integral around $\gamma_{P,r}$ would be zero by Proposition 1.4, and we know that this integral is 1. Let ω be any closed 1-form on U, and let $c = \int_{\gamma_{P,r}} \omega$. To show that $[\omega_P]$ spans H^1U, it suffices to show that $\omega - c \cdot \omega_P$ is exact, for then $[\omega] = c \cdot [\omega_P]$ in H^1U. Now $\omega - c \cdot \omega_P$ is a closed 1-form on U whose integral along γ is 0, and Lemma 1.17 implies that such a form is exact.

For (c), fix a positive number r less than the distance between P and Q. To see that $[\omega_P]$ and $[\omega_Q]$ are linearly independent, we must show that $a \cdot \omega_P + b \cdot \omega_Q$ is not exact unless a and b are both zero. The integral of this form along $\gamma_{P,r}$ is a, and the integral along $\gamma_{Q,r}$ is b, and by Proposition 1.4, these integrals must vanish if the form is exact. To show that $[\omega_P]$ and $[\omega_Q]$ span H^1U, let ω be any closed 1-form on U. Let $a = \int_{\gamma_{P,r}} \omega$ and let $b = \int_{\gamma_{Q,r}} \omega$. To complete the proof of (c), we must show that $\omega - a \cdot \omega_P - b \cdot \omega_Q$ is exact on U, for then $[\omega] = a \cdot [\omega_P] + b \cdot [\omega_Q]$ in H^1U. Now $\omega - a \cdot \omega_P - b \cdot \omega_Q$ is a closed 1-form on U whose integrals along $\gamma_{P,r}$ and $\gamma_{Q,r}$ both vanish, and an appeal to Lemma 1.18 completes the proof. □

Problem 5.2. Generalize the proposition from one and two to n points.

There is one other result we will need later in this chapter:

Proposition 5.3. *If A is a connected closed subset of $\mathbb{R}^2$, and P and Q are points in A, then $[\omega_P] = [\omega_Q]$ in $H^1(\mathbb{R}^2 \setminus A)$.*

Proof. In order to show that $\omega = \omega_P - \omega_Q$ is exact on $\mathbb{R}^2 \setminus A$ it suffices by Proposition 1.8 to show that $\int_\gamma \omega = 0$ whenever γ is a segmented closed path in $\mathbb{R}^2 \setminus A$. But

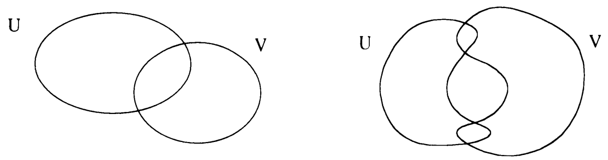
$$\int_\gamma \omega = \int_\gamma \omega_P - \int_\gamma \omega_Q = W(\gamma, P) - W(\gamma, Q).$$

Now Proposition 3.16 implies that the winding numbers $W(\gamma, P)$ and $W(\gamma, Q)$ are equal, since P and Q belong to the same connected component of $\mathbb{R}^2 \setminus \mathrm{Supp}(\gamma)$. □

Exercise 5.4. With A as in the proposition, and P in A, show that $[\omega_P] = 0$ in $H^1(\mathbb{R}^2 \setminus A)$ if and only if A is unbounded.

5b. The Coboundary Map

Our basic tool for studying H^0U and H^1U, and hence the topology of U, will be a method to study the relations among these groups for open sets U, V and their union $U \cup V$ and intersection $U \cap V$.



That there might be some relation can be seen already in Lemma 1.14. This showed, for example, that if $U \cap V$ is connected, i.e., $H^0(U \cap V)$ has dimension at most one, and if $H^1U = 0$, and $H^1V = 0$, then $H^1(U \cup V)$ must vanish as well. We want to generalize this to more complicated open sets.

For any two open sets U and V in the plane, we will define a linear map

$$\delta\colon H^0(U \cap V) \to H^1(U \cup V).$$

To do this, we will need the following basic fact:

Lemma 5.5. *Given open sets U and V in $\mathbb{R}^2$, any $\mathscr{C}^\infty$ function on*

$U \cap V$ can be written as the difference of two $\mathscr{C}^\infty$ functions, one that extends to a $\mathscr{C}^\infty$ function on U, the other to a $\mathscr{C}^\infty$ function on V.

Proof. This is a consequence of the existence of a partition of unity, for the simple (but not so trivial) case of the covering of $U \cup V$ by the open sets U and V. This says that there are $\mathscr{C}^\infty$ functions φ and ψ on $U \cup V$ such that $\varphi + \psi \equiv 1$ on $U \cup V$, such that the closure (in $U \cup V$) of the support of φ is contained in U, and the closure (in $U \cup V$) of the support of ψ is contained in V. (The support of a function is the set where it is not zero.) For the proof, see Appendix B2.

To use it to prove the lemma, given a $\mathscr{C}^\infty$ function f on $U \cap V$, define f_1 on U by the rule

$$f_1 = \begin{cases} \psi \cdot f & \text{on } U \cap V, \\ 0 & \text{on } U \setminus U \cap V. \end{cases}$$

The assumption on the support of ψ insures that any point in the set $U \setminus U \cap V$ has a neighborhood disjoint from the support of ψ, from which it follows that f_1 is $\mathscr{C}^\infty$ on all of U. Similarly, define a $\mathscr{C}^\infty$ function f_2 on V by

$$f_2 = \begin{cases} -\varphi \cdot f & \text{on } U \cap V, \\ 0 & \text{on } V \setminus U \cap V. \end{cases}$$

Then $f_1 - f_2 = (\psi + \varphi)f = f$ on $U \cap V$, as required. □

Construction of δ: $H^0(U \cap V) \to H^1(U \cup V)$. Given a locally constant function f on $U \cap V$, use the lemma to find $\mathscr{C}^\infty$ functions f_1 and f_2 on U and V respectively so that $f = f_1 - f_2$ on $U \cap V$. Since f is locally constant,

$$df_1 - df_2 = d(f_1 - f_2) = df = 0 \quad \text{on } U \cap V.$$

This means df_1 and df_2 agree on $U \cap V$, so there is a unique 1-form ω on $U \cup V$ that agrees with df_1 on U and with df_2 on V. This 1-form ω is closed, since it is even exact on each of U and V.

We define $\delta(f)$ to be the equivalence class in $H^1(U \cup V)$ determined by this closed form ω, i.e., $\delta(f) = [\omega]$. For this to be well defined, we must see how ω depends on the way we write f as the difference $f_1 - f_2$. We claim that a different choice would lead to a closed 1-form that differs from ω by an exact 1-form. To see this, suppose f_1' and f_2' were another choice of functions on U and V with $f_1' - f_2' = f$ on $U \cap V$. (These primes have nothing to do with derivatives!) Let ω' be the 1-form on $U \cap V$ that is df_1' on U and is df_2'

on V. Now since $f_1' - f_2' = f_1 - f_2$ on $U \cap V$,

$$f_1' - f_1 = f_2' - f_2$$

on $U \cap V$, so there is a $\mathscr{C}^\infty$ function g on $U \cup V$ that is $f_1' - f_1$ on U and is $f_2' - f_2$ on V. Then $dg = \omega' - \omega$, as required.

Lemma 5.6. *The mapping δ is a linear mapping of vector spaces, i.e., $\delta(f+g) = \delta(f) + \delta(g)$ and $\delta(c \cdot f) = c \cdot \delta(f)$, for f, g locally constant functions, and c a constant.*

Proof. This is just a matter of making the choices "linearly," and is better (and probably easier) to check for yourself than to read. Write $f = f_1 - f_2$ and $g = g_1 - g_2$ as in the construction of $\delta(f)$ and $\delta(g)$. Then $f + g = (f_1 + g_1) - (f_2 + g_2)$. If ω_f is the 1-form that is df_1 on U and df_2 on V, and ω_g is the 1-form that is dg_1 on U and dg_2 on V, then $\omega_f + \omega_g$ is the 1-form that is $d(f_1 + g_1)$ on U and $d(f_2 + g_2)$ on V. Therefore $\delta(f+g)$ is represented by the 1-form $\omega_f + \omega_g$, which by definition represents the sum $\delta(f) + \delta(g)$. This proves that δ preserves sums. The proof that it preserves multiplication by a scalar is similar, and left as an exercise. □

This map δ is called the *coboundary* map. In order to use it to compare $U \cap V$ with $U \cup V$, we need a description of its kernel and its image. The following two propositions do this.

Proposition 5.7. *A locally constant function f on $U \cap V$ is in the kernel of δ if and only if there are locally constant functions f_1 on U and f_2 on V so that $f = f_1 - f_2$ on $U \cap V$. In particular, if U and V are connected, then the kernel of δ consists of the constant functions on $U \cap V$.*

Proof. If $f = f_1 - f_2$ with f_1 and f_2 locally constant on U and V, these can be chosen for the construction of $\delta(f)$, and the corresponding form ω is zero. Conversely, if $\delta(f)$ is zero, the form ω of the construction from an equation $f = f_1 - f_2$ must be exact. Write $\omega = dg$. Then $df_1 = dg$ on U, and $df_2 = dg$ on V. This means that $f_1 - g$ is locally constant on U, and $f_2 - g$ is locally constant on V. And

$$f = f_1 - f_2 = (f_1 - g) - (f_2 - g)$$

is the difference of two such functions.

If U and V are connected, locally constant functions on them must be constant, and since f is the difference of two such functions, f is also constant. □

Exercise 5.8. Show that if U and V are connected, and $H^1(U \cup V) = 0$, then $U \cap V$ is also connected.

Proposition 5.9. *The class* $[\omega]$ *of a closed* 1*-form* ω *on* $U \cup V$ *is in the image of* δ *if and only if the restrictions of* ω *to* U *and to* V *are exact. In particular, if* $H^1U = 0$ *and* $H^1V = 0$, *then* δ *is surjective.*

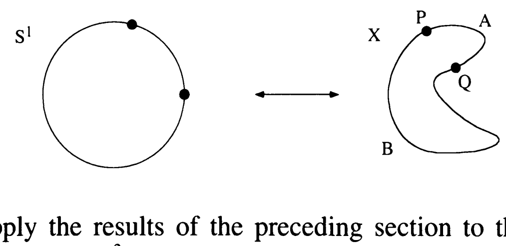
Proof. By construction, if ω is in the image of δ, there are functions f_1 on U and f_2 on V with $\omega = df_1$ on U and $\omega = df_2$ on V. Conversely, if there are such functions, the function $f = f_1 - f_2$ on $U \cap V$ is locally constant, since $df = df_1 - df_2 = \omega - \omega = 0$ on $U \cap V$, and $[\omega] = \delta(f)$ by construction. □

5c. The Jordan Curve Theorem

If X is a closed subset of the plane, there is a close relation between the topology of X and the topology of its complement. We prove some important cases of this fact here.

Theorem 5.10 (Jordan Curve Theorem). *If* $X \subset \mathbb{R}^2$ *is homeomorphic to a circle, then its complement* $\mathbb{R}^2 \setminus X$ *has two connected components, one bounded, the other unbounded. Any neighborhood of any point on* X *meets both of these components.*

Proof. Let P and Q be any two points in X. By considering a homeomorphism of X with the circle, we can write X as a union of two subsets A and B, each homeomorphic to a closed interval, with $A \cap B = \{P, Q\}$.



We will apply the results of the preceding section to the open sets $U = \mathbb{R}^2 \setminus A$ and $V = \mathbb{R}^2 \setminus B$. Note that

$$U \cup V = \mathbb{R}^2 \setminus \{P, Q\} \quad \text{and} \quad U \cap V = \mathbb{R}^2 \setminus X.$$

To show that $\mathbb{R}^2 \setminus X$ has two components, we want to show that the dimension of $H^0(U \cap V) = H^0(\mathbb{R}^2 \setminus X)$ is 2.

We know from Proposition 5.1(c) that $H^1(U \cup V) = H^1(\mathbb{R}^2 \setminus \{P, Q\})$ is a vector space of dimension 2, with a basis the classes of ω_P and ω_Q. Each of U and V is the complement of a subset homeomorphic to an interval. We will need an analogue of the Jordan curve theorem, but where the circle is replaced by an interval:

Theorem 5.11. *If $Y \subset \mathbb{R}^2$ is homeomorphic to a closed interval, then $\mathbb{R}^2 \setminus Y$ is connected.*

We postpone a discussion of this theorem to the end of this section, and show now how to use it to prove the Jordan curve theorem. Consider the boundary map

$$\delta\colon\ H^0(U \cap V) = H^0(\mathbb{R}^2 \setminus X) \quad \to \quad H^1(U \cup V) = H^1(\mathbb{R}^2 \setminus \{P, Q\}).$$

Our goal is to show that the image and kernel of δ are both one dimensional, which will imply by the rank-nullity theorem (see Appendix C) that $H^0(\mathbb{R}^2 \setminus X)$ is two dimensional, which means that $\mathbb{R}^2 \setminus X$ has two connected components.

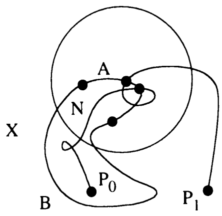
Since, by Theorem 5.11, U and V are connected, we can apply Proposition 5.7, so the kernel of δ consists of the constant functions on $\mathbb{R}^2 \setminus X$, which is therefore one dimensional. We claim that the image of δ consists of those linear combinations $a \cdot [\omega_P] + b \cdot [\omega_Q]$ of the basis elements that have $a + b = 0$. This means that $[\omega_P] - [\omega_Q]$ forms a basis for the image of δ, and completes the proof of the claim that the image is one dimensional.

By Proposition 5.9, the image of δ consists of those linear combinations $a \cdot [\omega_P] + b \cdot [\omega_Q]$ such that the restrictions of $a \cdot \omega_P + b \cdot \omega_Q$ to U and to V are exact. Since P and Q are in the same connected component of $\mathbb{R}^2 \setminus U$ and of $\mathbb{R}^2 \setminus V$, it follows from Proposition 5.3 that $\omega_P - \omega_Q$ is exact on U and on V. If $a + b = 0$, it follows that $a \cdot \omega_P + b \cdot \omega_Q = a \cdot (\omega_P - \omega_Q)$ is exact on U and on V, so $a \cdot [\omega_P] + b \cdot [\omega_Q]$ is in the image of δ if $a + b = 0$. Conversely, suppose the restriction of $\omega = a \cdot \omega_P + b \cdot \omega_Q$ to U (and to V) is exact. Let $\gamma = \gamma_{0,r}$ be a circle about the origin, with r so large that X is contained inside this circle. Since ω is exact on U, Proposition 1.4 guarantees that $\int_\gamma \omega = 0$. Since P and Q are inside the circle, $\int_\gamma \omega_P = 1$ and $\int_\gamma \omega_Q = 1$, so

$$0 = \int_\gamma \omega = a \cdot \int_\gamma \omega_P + b \cdot \int_\gamma \omega_Q = a + b.$$

This completes the proof that $\mathbb{R}^2 \setminus X$ has two connected components. Note that since X is bounded, one of these components must contain everything outside some large disk; this is the unbounded

component, while the other must be bounded. To verify the last assertion of the theorem, suppose N is a neighborhood of some point in X. We may divide X into two pieces, as in the preceding discussion, so that one of them, say A, lies entirely in N. Take two points, say P_0 and P_1, one in each of the two components of $\mathbb{R}^2 \setminus X$.

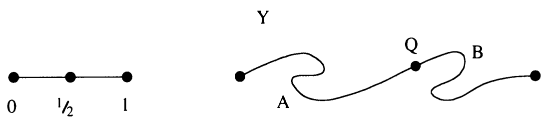


By Theorem 5.11, $\mathbb{R}^2 \setminus B$ is connected, so there is a path $\gamma(t)$, $0 \le t \le 1$, from P_0 to P_1 in $\mathbb{R}^2 \setminus B$. This path must hit A, for if it didn't, it would connect the two points in $\mathbb{R}^2 \setminus X$. By looking at the first and last time the path hits A, i.e., the minimal and maximal t such that $\gamma(t)$ is in A (which is a closed set in $[0, 1]$), it follows that N must contain points of both components: for t near to but less than the minimum, $\gamma(t)$ will be in N and the component of P_0, and for t near to but greater than the maximum, $\gamma(t)$ will be in N and the other component. □

Remark 5.12. (1) It follows in particular that X is nowhere dense: no point has a neighborhood contained in X. Note that an arbitrary continuous image of a circle or interval need not have this property.

(2) The fact that points on X are "accessible" from both sides is strengthened in Problem 10.24.

***Proof of Theorem* 5.11.** The proof of Theorem 5.11 uses the same ideas, but with one new wrinkle. Choose a homeomorphism from the interval $[0, 1]$ to Y. Let A be the subset of Y corresponding to the left half of the interval $[0, {}^1\!/_2]$, and B the subset corresponding to the right half $[{}^1\!/_2, 1]$, and $Q = A \cap B$ the point corresponding to ${}^1\!/_2$.



We apply the basic construction to $U = \mathbb{R}^2 \setminus A$ and $V = \mathbb{R}^2 \setminus B$, with $U \cap V = \mathbb{R}^2 \setminus Y$ and $U \cup V = \mathbb{R}^2 \setminus \{Q\}$. It might seem that little will be gained by this, since each of U, V, and $U \cap V$ is of the same form: the complement of a subset homeomorphic to an interval. But we will see that some progress has been made.

Suppose $\mathbb{R}^2 \setminus Y$ is not connected, and let P_0 and P_1 be points in two different connected components. The claim we will prove is: *P_0 and P_1 must be in different connected components of $\mathbb{R}^2 \setminus A$ or in different connected components of $\mathbb{R}^2 \setminus B$* (or both). We use the map

$$\delta\colon\ H^0(U \cap V) = H^0(\mathbb{R}^2 \setminus Y) \ \to\ H^1(U \cup V) = H^1(\mathbb{R}^2 \setminus \{Q\}).$$

Now $H^1(\mathbb{R}^2 \setminus \{Q\})$ is generated by the class $[\omega_Q]$ of ω_Q. By Proposition 5.9 the image of δ consists of classes $a \cdot [\omega_Q]$ such that $a \cdot \omega_Q$ is exact on U and V. If $a \cdot \omega_Q$ is exact on U, the integral of $a \cdot \omega_Q$ around a large circle is zero, which implies that $a = 0$. So the image of δ is zero. In other words, every locally constant function on $\mathbb{R}^2 \setminus Y$ is in the kernel of δ.

By Proposition 5.7, a function f in the kernel of δ must have the form $f_1 - f_2$, where f_1 and f_2 are locally constant functions on U and on V, respectively. Take f to be any locally constant function on $U \cap V$ that takes on different values at the two points P_0 and P_1, which is possible since they are in different components. If the claim were false, and P_0 and P_1 were in the same component of U and of V, then each of f_1 and f_2 would have to take on the same values at P_0 and P_1. But then their difference f would also take on the same values, which is a contradiction.

The progress made by proving this claim comes from the fact that the two sets A and B are *smaller* than Y. To take advantage of this, we can argue as follows: Let Y_1 be one of the halves of Y such that P_0 and P_1 are in different components of $\mathbb{R}^2 \setminus Y_1$. Repeat the argument, cutting Y_1 into two pieces (corresponding to cutting the half interval $[0, {}^1\!/_2]$ or $[{}^1\!/_2, 1]$ into equal pieces). For one of the two pieces, say Y_2, by the same argument, the two points P_0 and P_1 are still in different components of $\mathbb{R}^2 \setminus Y_2$. Continuing in this way, we get a nested sequence of subsets

$$Y \supset Y_1 \supset Y_2 \supset Y_3 \supset \ldots \supset Y_n \supset \ldots$$

with the property that P_0 and P_1 are in different components of $\mathbb{R}^2 \setminus Y_n$ for all n, with the intersection of all these subsets Y_n being a single point P in Y.

Since the complement of a point is connected, there is a path from P_0 to P_1 in $\mathbb{R}^2$ that doesn't pass through P. Some neighborhood N of

P is disjoint from this path, and, for large n, Y_n is contained in N. But this forces P_0 and P_1 to be in the same component of $\mathbb{R}^2 \setminus Y_n$, a contradiction. This finishes the proof of Theorem 5.11, and hence of the full Jordan curve theorem. □

5d. Applications and Variations

The same ideas can be used to calculate the number of connected components of the complements of many other subsets in the plane. For many of these variations, we need the following generalization of Proposition 5.1, which will be proved in Chapter 9 (see §9c and Lemma 9.1).

> (*) Let K be a compact, nonempty subset of the plane. (a) If K is connected, then $H^1(\mathbb{R}^2 \setminus K)$ is one-dimensional, generated by $[\omega_P]$ for any $P \in K$. (b) If K is not connected, and P and Q are in different components of K, then $[\omega_P]$ and $[\omega_Q]$ are linearly independent in $H^1(\mathbb{R}^2 \setminus K)$; if K has exactly two connected components, then $[\omega_P]$ and $[\omega_Q]$ form a basis for $H^1(\mathbb{R}^2 \setminus K)$.

Exercise 5.13. Show that, if A and B are compact connected subsets in the plane such that $A \cap B$ is not connected (and not empty), then $\mathbb{R}^2 \setminus (A \cup B)$ is not connected.

Exercise 5.14. Show that if Y is a subset of the plane homeomorphic to a closed rectangle or a closed disk, then the complement is connected.

Exercise 5.15. Show that if X is a subset of the plane homeomorphic to a closed annulus, then the complement has two connected components.

Exercise 5.16. Show that if X is a subset of the plane homeomorphic to a figure 8, or a "theta" Θ, then the complement has three connected components.

Here is a simple application of the Jordan curve theorem. Let D be a closed disk, D° its interior, and C its boundary circle.

Proposition 5.17. *Let $f: D \to \mathbb{R}^2$ be a continuous, one-to-one mapping. Then $\mathbb{R}^2 \setminus f(C)$ has two connected components, which are*

$$f(D^\circ) \quad \text{and} \quad \mathbb{R}^2 \setminus f(D).$$

In particular, $f(D^\circ)$ is an open subset of the plane.

Proof. Recall that a continuous, one-to-one mapping on a compact set such as D or C must be a homeomorphism onto its image. So the Jordan curve theorem applies to the image of C, and its complement has two connected components. The first displayed set is connected since it is the continuous image of a connected space, and the second is connected by Exercise 5.14. They are disjoint, and their union is $\mathbb{R}^2 \setminus f(C)$. It follows immediately that they must be the two components of $\mathbb{R}^2 \setminus f(C)$. In particular, since the components of an open subset of the plane are open, it follows that $f(D^\circ)$ is open. □

The following is another intuitively "obvious" result that is not so easy to prove by hand (try it!):

Corollary 5.18 (Invariance of Domain). *If U is an open set in the plane, and $F: U \to \mathbb{R}^2$ is a continuous, one-to-one mapping, then $F(U)$ is an open subset of $\mathbb{R}^2$, and F is a homeomorphism of U onto $F(U)$.*

Proof. Take any P in U, and a closed disk D containing P and contained in U. By the proposition, the image of the interior of the disk must be open. This gives an open neighborhood of the image point in $F(U)$, which implies that $F(U)$ is open, and that F is a homeomorphism of U with $F(U)$. □

In particular, if two subsets of the plane are homeomorphic, and one is an open subset, the other must also be open. So if one is a domain (a connected open subset), the other must be as well.

Another application is another proof of a result from Chapter 4.

Corollary 5.19. *There is no subset of the plane that is homeomorphic to a two-sphere.*

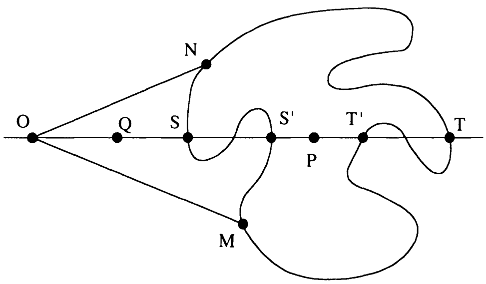
Proof. Suppose $f: S \to \mathbb{R}^2$ maps the sphere homeomorphically onto a subset X of $\mathbb{R}^2$. Let A and B be the upper and lower closed hemispheres, with intersection C, and let $A^\circ = A \setminus C$ and $B^\circ = B \setminus C$. Note that A is homeomorphic to a disk, with C corresponding to the boundary. The complement of the image $f(C)$ of the circle in the plane has two connected components, and by the proposition (applied to the map from A to the plane), $f(A^\circ)$ must be one of them. Since $f(A^\circ)$ is

contained in $f(A)$, it must be the bounded component. The same reasoning applies to the map from B, so $f(B^\circ)$ must also be the bounded component of the complement of $f(C)$. But then $f(A^\circ) = f(B^\circ)$, which contradicts the fact that f is one-to-one. □

This section concludes with a few related results, mostly in the form of exercises and problems, which can be sampled according to your interest or perseverance. The next proposition verifies that a Jordan curve winds once around each point inside. (This result will not be used elsewhere.)

Proposition 5.20. *Let C be a circle, $F: C \to \mathbb{R}^2$ a one-to-one, continuous mapping. Then $W(F,P) = \pm 1$ for P in the bounded component of $\mathbb{R}^2 \setminus F(C)$.*

Proof. In the course of the proof, we will make several assertions about winding numbers, leaving their proofs as exercises in using the properties proved in Chapter 3. The mapping F is a homeomorphism onto its image $X = F(C)$. By the Jordan curve theorem, the complement has two components, and we know the winding number is 0 for points in the unbounded component, and it is constant in the bounded component. So it suffices to find one point P with $W(F,P) = \pm 1$. Take a horizontal line L that has points of X on both sides of it. Take a point O on L so that X lies in the half plane to the right of O. Let S and T be the nearest and farthest points from O in $X \cap L$, and let M and N be points on X below and above L such that no other points of X lie on the line segments from O to M and from O to N.



The points M and N correspond by F to two points m and n of the circle C. Let A be the image by F of the counterclockwise arc from m to n, and let B be the image of the other clockwise arc from m to n. Let γ_A be a closed path starting at O, going along the segment to M, then traveling along A to N, and back to O along the segment;

define γ_B similarly using B in place of A. We choose these paths along A and B by using the mapping F, so that

$$W(\gamma_A, P) - W(\gamma_B, P) = W(F, P)$$

for all P not in X or the line segments from O to M or N.

The answer will depend on whether the point T is in A or in B. Suppose first that T is in A. Then $W(\gamma_B, T) = 0$, so $W(\gamma_B, T') = 0$ for all T' in $A \cap L$. For a point Q on the line between O and S, $W(\gamma_B, Q) = 1$. It follows that S is not in A, so it must be in B. Since $W(\gamma_A, Q) = 1$, we must have $W(\gamma_A, S) = 1$, and therefore $W(\gamma_A, S') = 1$ for all S' in $B \cap L$. Let S' be the point in $B \cap L$ that is farthest from O, and let T' be the point in $A \cap L$ that is closest to O, and let P be any point on L between S' and T'. Then

$$W(F, P) = W(\gamma_A, P) - W(\gamma_B, P) = W(\gamma_A, S') - W(\gamma_B, T') = 1.$$

If T is in B, a similar argument shows that $W(F, P) = -1$. □

The proof gives a criterion to tell whether the winding number is $+1$ or -1: it is $+1$ when four points M, N, S, and T chosen as in the proof have the same relative position as the four corresponding points of the circle. (This proof, from Ahlfors (1979), also shows directly, without the Jordan curve theorem, that the complement of X has at least two components.)

It is a fact that if X is a subset of $\mathbb{R}^2$ homeomorphic to a circle, then the bounded component of the complement is homeomorphic to an open disk, and the unbounded component is homeomorphic to a disk minus a point. In fact, the Riemann mapping theorem of complex analysis implies that any connected plane domain with $H^1U = 0$ is analytically isomorphic to (in particular, diffeomorphic to) an open disk. Even more is true: any homeomorphism of S^1 with X can be extended to a homeomorphism from $\mathbb{R}^2$ onto $\mathbb{R}^2$, mapping the inside and outside of the circle to the two components of $\mathbb{R}^2 \setminus X$. This last is called the Schoenflies theorem. For elementary (but not so simple) proofs, see Newman (1939). This is a case where the analogue in higher dimensions is more complicated: a homeomorphism of S^2 in $\mathbb{R}^3$, or of S^{n-1} in $\mathbb{R}^n$ for $n \geq 3$, will separate the space into two connected components, as we will see in Chapter 23, but the inside need not be homeomorphic to an open disk; for a picture of a counterexample, "Alexander's horned sphere," see Hocking and Young (1988). However, if the embedding extends to an embedding of a product of S^{n-1} with an interval into $\mathbb{R}^n$, then this wild behavior cannot occur (see Bredon (1993), §19).

Problem 5.21. Prove the following strong form of Euler's theorem. Let X be a subset of the plane that is a union of $v \geq 1$ points and $e \geq 0$ edges. The edges are assumed to be images of continuous maps from $[0, 1]$ to $\mathbb{R}^2$, each of which maps 0 and 1 into the set of vertices, and maps the open interval $(0, 1)$ one-to-one into the complement of the set of vertices. In addition, these open edges are assumed to be disjoint. Suppose X has k connected components. Show that $\mathbb{R}^2 \setminus X$ has $f = e - v + k + 1$ connected components, i.e.,

$$v - e + f = 2 + (k - 1).$$

You may enjoy comparing your argument in the last problem with that given in Rademacher and Toeplitz (1957), §12, as well as the applications given there to the problem of coloring maps. Can you spot where they make implicit assumptions amounting to what we proved in this chapter?

Problem 5.22. Show that the following graphs cannot be embedded in the plane: (i) the graph with vertices $P_1, P_2, P_3, Q_1, Q_2, Q_3$, with an edge between each P_i and each Q_j; and (ii) the graph with five vertices, and an edge between each pair of distinct vertices. (It is a theorem of Kuratowski that any finite graph not containing a subgraph homeomorphic to one of these two examples *can* be embedded in the plane.)

Problem 5.23. Let U be any connected open set in the plane. (a) Show that if $X \subset U$ is homeomorphic to a closed interval, then $U \setminus X$ is connected. (b) Show that if $X \subset U$ is homeomorphic to a circle, then $U \setminus X$ has two connected components. (c) If $X \subset U$ is a graph as in Problem 5.1, show that $U \setminus X$ has $e - v + k + 1$ connected components.

Exercise 5.24. Show that Theorem 5.11, the Jordan curve theorem (without mention of bounded or unbounded components), and the result of Problem 5.21 remain valid when $\mathbb{R}^2$ is replaced by a sphere.

Problem 5.25. Find two graphs in a sphere that are homeomorphic, but such that there is no homeomorphism of the sphere taking one onto the other.

Problem 5.26. Show that there is no one-to-one continuous map from a Moebius band into the plane.

Problem 5.27. Suppose X is a subset of the plane homeomorphic to a circle, and P_1 and P_2 are points in the complement that are joined by a path that crosses X n times. Show that P_1 and P_2 are in the same component of the complement if n is even, and the opposite component if n is odd. (A complete answer should include a precise definition of what it means for a path to cross X at a point!)

It is again an excellent project to speculate on the higher-dimensional generalizations of the results of this chapter. For example, if you assume the fact that if X is any subset of $\mathbb{R}^n$ homeomorphic to a sphere S^{n-1} (resp. a ball D^n), then $\mathbb{R}^n \setminus X$ has two (resp. one) connected components, can you state and prove the invariance of domain for open sets in $\mathbb{R}^n$?

Problem 5.28. A topological *surface with boundary* is defined to be a Hausdorff space such that every point P has a neighborhood homeomorphic either to the open disk $D^\circ = \{(x, y): x^2 + y^2 < 1\}$ or the half disk $\{(x, y) \in D^\circ: y \geq 0\}$, with P corresponding to the origin; P is an *interior* or *boundary* point according to which case occurs. (a) Show that this notion is well defined: a point cannot be both an interior point and a boundary point. (b) Show that homeomorphic surfaces have homeomorphic boundaries (so the Moebius band is not homeomorphic to a cylinder).

CHAPTER 6

Homology

6a. Chains, Cycles, and H_0U

As we have seen, it frequently happens that one wants to compute winding numbers or integrals along a succession of paths, counting some positively and some negatively. For example, the integral around the boundary of a rectangle is the sum of integrals over two of its sides, minus the sum of the integrals over the other two sides. In this section we formalize these ideas, by introducing the notion of a 1-*chain*. A 1-chain γ in U is an expression of the form

$$\gamma = n_1\gamma_1 + n_2\gamma_2 + \ldots + n_r\gamma_r,$$

where each γ_i is a continuous path in U, and each n_i is an integer. For simplicity, so we will not have to mention the interval each path might be defined on, we will take all paths from now on to be defined on the unit interval $[0, 1]$. The paths that are *constant*, that is, that map $[0, 1]$ to one point of U, can be ignored in the present story. For example, their winding numbers are zero, all integrals over them are zero. We will agree that if we meet a sum $\Sigma n_i\gamma_i$ where some of the γ_i are constant paths, we simply throw away any constant paths that occur. Another way to say this is that we identify two expressions $\Sigma n_i\gamma_i$ and $\Sigma n_i'\gamma_i'$ if their difference $\Sigma n_i\gamma_i - \Sigma n_i'\gamma_i'$ is a linear combination of constant paths.

We will make this notion more precise in a moment, but for now we note that, whatever it means, it is clear how we should define the

winding number of a 1-chain γ with respect to P:

$$W(\gamma,P) = n_1W(\gamma_1,P) + \ldots + n_rW(\gamma_r,P);$$

this will be defined provided P is not in the support of any of the paths γ_i. In Chapter 9 we will also define the integral of a closed 1-form along a path, and it will extend additively in the same way to integrals over 1-chains. Two 1-chains should be regarded as the same when each path occurs with the same multiplicity in both 1-chains. A 1-chain will have a unique expression as shown, provided the paths γ_i are all taken to be distinct and nonconstant, and all the coefficients are taken to be nonzero. (There is also the *zero* 1-chain, written $\gamma = 0$, which has no paths at all.)

To make this precise, define a *1-chain* in U to be a function that assigns to every nonconstant path in U some integer, with the property that the function is zero for all but a finite number of paths. If $\gamma_1, \ldots, \gamma_r$ are the paths for which the value is not zero, and the value of this function on γ_i is n_i, we write the 1-chain as $n_1\gamma_1 + \ldots + n_r\gamma_r$. From this definition it is clear how to add and subtract 1-chains: one just adds or subtracts the corresponding values on each path, or the coefficients in such expressions. In this way the 1-chains form an abelian group,[4] with the operation in the group written additively. Any path γ is identified with the 1-chain $1 \cdot \gamma$, the corresponding function taking the value 1 on γ and 0 on all other paths.

In practice, we will not use this "functional" terminology, but just write 1-chains as formal linear combinations of paths. Either way, specifying a 1-chain is the same as specifying a finite number of paths, and assigning an integer to each. It should be emphasized that in this definition two paths are the same only if defined by exactly the same mapping. However, the following problem indicates some common variations that are possible.

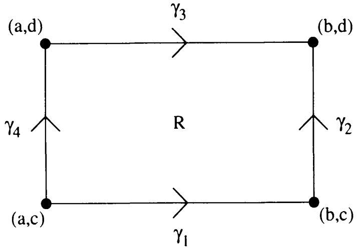
Problem 6.1. Call two paths equivalent if they differ by a monotone increasing reparametrization. Show that this is an equivalence relation. Show how to define an abelian group of 1-chains by using equivalence classes of paths instead of paths. Do the same if paths are called equivalent when they are homotopic with the same endpoints.

[4] If you are familiar with the language from algebra, the chains are elements of the free abelian group on the set of nonconstant paths. See Appendix C for more about free abelian groups.

Some particular paths and 1-chains will be important. For any two points P and Q, the *straight path* from P to Q will be the path

$$\gamma(t) \; = \; P + t(Q-P) = (1-t)P + tQ, \qquad 0 \le t \le 1 .$$

If R is a bounded rectangle with sides parallel to the axes, the *boundary* ∂R of R is the 1-chain $\gamma_1 + \gamma_2 - \gamma_3 - \gamma_4$, where each γ_i is the straight path shown:



If $R = [a,b] \times [c,d]$, then γ_1 is the straight path from (a,c) to (b,c), γ_2 from (b,c) to (b,d), γ_3 from (a,d) to (b,d), and γ_4 from (a,c) to (a,d).

If D is a disk of radius r around a point P, the *boundary* $\gamma = \partial D$ is the counterclockwise path around the circle

$$\gamma(t) \; = \; P + (r\cos(2\pi t), r\sin(2\pi t)), \qquad 0 \le t \le 1 .$$

We want to define the notion of a *closed* 1-*chain*, also called a 1-*cycle*. This should mean that each point occurs as many times as an initial point as it does as a final point, counting multiplicities correctly. For example, any closed path, such as the preceding ∂D, is a closed 1-chain, and the boundary ∂R of a rectangle is closed. The precise definition is as follows. For $\gamma = n_1\gamma_1 + \ldots + n_r\gamma_r$, if γ_i is a path from P_i to Q_i, we define γ to be a *closed* 1-*chain* if for each point T occurring as a starting or ending point of any γ_i,

$$\sum_{P_i = T} n_i \; = \; \sum_{Q_j = T} n_j .$$

Exercise 6.2. Verify that the sum or difference of closed 1-chains is closed.

Define a 0-*chain* in U to be a formal finite linear combination $m_1P_1 + \ldots + m_sP_s$ of points in U, with integer coefficients. Precisely, it is a function from U to $\mathbb{Z}$ that is zero outside a finite set (or

an element of the free abelian group on the set of points of U). In practice it means to pick out a finite set of points and assign a positive or negative multiplicity to each. For any 1-chain $\gamma = n_1\gamma_1 + \ldots + n_r\gamma_r$, where γ_i: $[0, 1] \to U$ is a path, define the *boundary* $\partial\gamma$ of γ to be the 0-chain

$$\partial\gamma = n_1(\gamma_1(1) - \gamma_1(0)) + \ldots + n_r(\gamma_r(1) - \gamma_r(0)).$$

The preceding definition of a closed 1-chain can be said simply in this language: a 1-chain is closed exactly when its boundary is zero.

Let Z_0U be the group of 0-chains. A 0-chain ζ is called a *0-boundary* if there is a 1-chain γ such that $\zeta = \partial\gamma$. These 0-boundaries form a subgroup of Z_0U which is denoted by B_0U. For example, if P and Q are in the same component of U, then the 0-chain $Q - P$ is in B_0U, since it is the boundary of any path from P to Q. The quotient group Z_0U/B_0U is called the *0th homology group* of U, and is denoted H_0U:

$$H_0U = Z_0U/B_0U.$$

We will see that although the groups Z_0U and B_0U are large (even uncountable), the quotient group is small: it simply measures how many connected components U has:

Proposition 6.3. *The group H_0U is canonically isomorphic to the free abelian group on the set of path-connected components of U.*

Proof. Let F be the free abelian group on the set of path-connected components of U. The map that takes a point to the path component containing it determines a surjective homomorphism from Z_0U to F. We claim that the kernel of this homomorphism is exactly the group of boundaries B_0U. This will conclude the proof, since such a homomorphism determines a canonical isomorphism of Z_0U/B_0U with F (see Appendix C). Any boundary is in the kernel, since the endpoints of a path must be in the same component. Conversely, if a 0-cycle is in the kernel, the total of the coefficients appearing in front of points in any given component must be zero. Such a 0-cycle can be written (not necessarily uniquely) in the form $\Sigma(Q_i - P_i)$, where, in each term, P_i and Q_i are in the same component. As we saw before the proof, such a 0-cycle is a boundary. □

We will use this proposition to determine the number of connected components of U, by finding other ways of calculating H_0U. If we show that H_0U has rank n, we will know that U has exactly n connected components. This depends on the algebraic fact that a free abelian group has a well-defined rank; this is proved in Appendix C.

(Alternatively, one could replace all the integer coefficients in all our 0-chains and 1-chains by real numbers. Then we would find that H_0U is a real vector space of dimension n, where n is the number of components, and appeal to the fact that a vector space has a well-defined dimension.)

There is a homomorphism from H_0U to the integers $\mathbb{Z}$, defined by the map that takes each connected component of U to 1. In other words, it takes the class of a 0-cycle $\zeta = \Sigma n_i P_i$ to the sum Σn_i of the coefficients. This is called the *degree* homomorphism. It is an isomorphism exactly when U is connected.

6b. Boundaries, H_1U, and Winding Numbers

The group of 1-chains on U is denoted C_1U. The subgroup of closed 1-chains, or 1-cycles, is denoted Z_1U. There are some closed 1-chains in U, called 1-*boundaries*, that play a particularly simple role. They will turn out to be exactly those 1-chains for which winding numbers around points not in U vanish, and for which all integrals of closed 1-forms in U also vanish. These come from boundaries of continuous mappings Γ from a square $R = [0, 1] \times [0, 1]$ into U. For such a mapping, define the 1-chain $\partial\Gamma$ by the formula

$$\partial\Gamma \;=\; \gamma_1 + \gamma_2 - \gamma_3 - \gamma_4,$$

where γ_1, γ_2, γ_3, and γ_4 are the paths obtained by restricting Γ to the four sides of the square, as in §3b. (We should note here that one or more of these four paths γ_i could be constant paths, in which case we omit them from the formula for $\partial\Gamma$.) We call a 1-chain γ a *boundary*, or a *boundary* 1-*chain*, or 1-*boundary*, in U if it can be written as a finite linear combination (with integer coefficients) of boundaries of such maps on rectangles. Two closed 1-chains in U are *homologous* if the difference between them is a boundary in U.

We will need to know that some other 1-chains are boundaries. The following lemma considers what happens when one reparametrizes, subdivides, or deforms a path.

Lemma 6.4. (*a*) *Let* γ: $[0, 1] \to U$ *be a path. Let* φ: $[0, 1] \to [0, 1]$ *be a continuous function. If* $\varphi(0) = 0$ *and* $\varphi(1) = 1$, *then* $\gamma - \gamma \circ \varphi$ *is a boundary in* U; *if* $\varphi(0) = 1$ *and* $\varphi(1) = 0$, *then* $\gamma + \gamma \circ \varphi$ *is a boundary.*

(*b*) *Let* γ: $[0, 1] \to U$ *be a path, let* $0 \le c \le 1$, *and let* σ *and* τ *be*

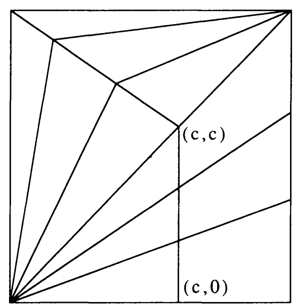
the restriction of γ *to* $[0,c]$ *and* $[c,1]$, *but scaled to be defined on the unit interval, i.e.*,

$$\sigma(t) = \gamma(c\cdot t), \quad 0 \le t \le 1; \qquad \tau(t) = \gamma(c+(1-c)\cdot t), \quad 0 \le t \le 1.$$

Then $\gamma - \sigma - \tau$ *is a boundary in* U.

(*c*) *If* γ *and* δ *are paths in* U *that are homotopic, either as paths with the same endpoints, or as closed paths, then* $\gamma - \delta$ *is a boundary in* U.

Proof. Part (a) was proved in the course of proving Corollary 3.9. Part (b) is trivial if $c = 0$ or $c = 1$. Otherwise, we construct a mapping Γ: $[0,1] \times [0,1] \to U$ as indicated in the following diagram:



A little scratch-work in analytic plane geometry produces the formula for Γ:

$$\Gamma(t,s) = \begin{cases} \gamma(t) & \text{if } s \le t, \\ \gamma(c\cdot s + (1-c)\cdot t) & \text{if } s \ge t. \end{cases}$$

Note that the two expressions for Γ agree where $s = t$, so they define a continuous mapping. Moreover, $\Gamma(t,0) = \gamma(t)$, $\Gamma(1,s) = \gamma(1)$, $\Gamma(t,1) = \tau(t)$, and $\Gamma(0,s) = \sigma(s)$, so $\partial\Gamma = \gamma - \tau - \sigma$, which proves (b). Part (c) follows directly from the definition. □

Exercise 6.5. Show that if U is starshaped, then every closed 1-chain on U is a 1-boundary.

The 1-boundaries form a subgroup, denoted B_1U, of the group Z_1U of 1-cycles. The quotient group Z_1U/B_1U is called the *first homology group* of U, and is denoted H_1U. One of our aims in this book will be to study and use this group. We will see that, as the 0th group

H_0U measures the simplest topological fact about U—how many connected components it has—the 1st group H_1U measures how many "holes" there are in U. Two closed 1-chains are homologous exactly when they have the same image in H_1U, in which case we say that they define the same *homology class*.

The support Supp(γ) of a 1-chain γ in the plane is the union of the supports of the paths that appear in γ with nonzero coefficients. If P is a point not in the support, we define the *winding number* of γ around P to be

$$W(\gamma, P) = n_1 W(\gamma_1, P) + \ldots + n_r W(\gamma_r, P),$$

where $\gamma = n_1\gamma_1 + n_2\gamma_2 + \ldots + n_r\gamma_r$.

Proposition 6.6. *If γ is a closed 1-chain, then, for any P not in the support of γ, $W(\gamma, P)$ is an integer.*

Proof. Let $\gamma = n_1\gamma_1 + \ldots + n_r\gamma_r$ be any 1-chain, with γ_i a path. For each point T that occurs as an endpoint of any of the paths γ_i, choose an angle ϑ_T for T with respect to P. (Such an angle is measured counterclockwise from a horizontal line to the right from P; it is determined only up to adding multiples of 2π.) If γ_i is a path from P_i to Q_i, then $W(\gamma_i, P) = (1/2\pi)(\vartheta_{Q_i} - \vartheta_{P_i}) + N_i$ for some integer N_i. Therefore,

$$\begin{aligned} W(\gamma, P) &= \sum_{i=1}^{r} n_i\left(\frac{1}{2\pi}(\vartheta_{Q_i} - \vartheta_{P_i}) + N_i\right) \\ &= \frac{1}{2\pi}\sum_{i=1}^{r} n_i(\vartheta_{Q_i} - \vartheta_{P_i}) + \sum_{i=1}^{r} n_i N_i. \end{aligned}$$

Suppose $\partial\gamma = \Sigma_{i=1}^{r} n_i(Q_i - P_i) = m_1T_1 + \ldots + m_sT_s$. Then we have

$$W(\gamma, P) = \frac{1}{2\pi}(m_1\vartheta_{T_1} + \ldots + m_s\vartheta_{T_s}) + \sum_{i=1}^{r} n_i N_i.$$

In particular, if γ is closed, i.e., $\partial\gamma = 0$, then the first sum vanishes, so $W(\gamma, P)$ is an integer. □

Lemma 6.7. *If γ is a 1-boundary in $\mathbb{R}^2 \setminus \{P\}$, then $W(\gamma, P) = 0$. If two 1-chains differ by a 1-boundary in $\mathbb{R}^2 \setminus \{P\}$, then they have the same winding number around P.*

Proof. If $\gamma = \partial\Gamma$ is the boundary of a map from $[0, 1] \times [0, 1]$ into $\mathbb{R}^2 \setminus \{P\}$, Theorem 3.6 implies that $W(\gamma, P) = 0$. For a general boundary $\gamma = \Sigma n_i(\partial\Gamma_i)$, $W(\gamma, P) = \Sigma n_i W(\partial\Gamma_i, P) = 0$. □

Proposition 6.8. *If γ is a closed 1-chain on $\mathbb{R}^2$, then the function $P \mapsto W(\gamma, P)$ is constant on connected components of $\mathbb{R}^2 \setminus Supp(\gamma)$, and vanishes on the unbounded component.*

Proof. To show that the function is locally constant, it suffices to show that it is constant on a disk D about a point P that does not meet the support of γ. We want to show that $W(\gamma, P) = W(\gamma, Q)$, with Q a point of D. Let v be the vector from P to Q. Let $\gamma = \Sigma_{i=1}^{r} n_i \gamma_i$, with each γ_i a path. We know from Exercise 3.4 that $W(\gamma, P) = W(\gamma + v, P + v) = W(\gamma + v, Q)$, where $\gamma + v$ is the 1-chain $\Sigma_{i=1}^{r} n_i(\gamma_i + v)$, so it is enough to show that

$$W(\gamma, Q) = W(\gamma + v, Q).$$

By the lemma, this follows if we verify that the difference of the 1-chains γ and $\gamma + v$ is a 1-boundary in $\mathbb{R}^2 \setminus \{Q\}$. Define mappings Γ_i from $[0, 1] \times [0, 1]$ to $\mathbb{R}^2 \setminus \{Q\}$ by the formula

$$\Gamma_i(t, s) = \gamma_i(t) + s \cdot v.$$

The boundary of Γ_i has the paths γ_i and $\gamma_i + v$ on the bottom and top, and straight line paths from endpoints of γ_i to their translations by v, along the sides. The fact that γ is closed means that these straight line paths from the sides cancel in the sum $\Sigma_{i=1}^{r} n_i(\partial \Gamma_i)$. Therefore,

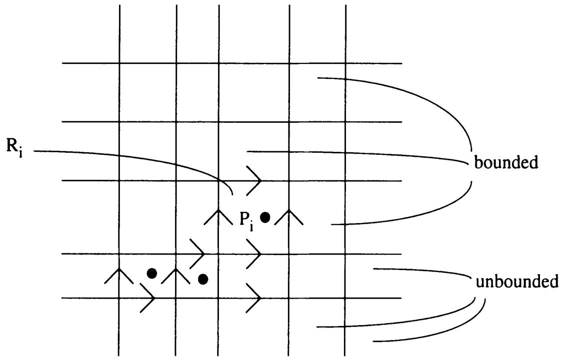
$$\sum_{i=1}^{r} n_i(\partial \Gamma_i) = \sum_{i=1}^{r} n_i \gamma_i - \sum_{i=1}^{r} n_i(\gamma_i + v) = \gamma - (\gamma + v).$$

This shows that $\gamma - (\gamma + v)$ is a boundary in $\mathbb{R}^2 \setminus \{Q\}$, and concludes the proof that the function is locally constant. On the unbounded component, we may take the point P far to the left of the support of γ, so that there is one angle function ϑ defined on all the paths occurring in γ. With $\gamma = \Sigma_{i=1}^{r} n_i \gamma_i$, and γ_i a path from P_i to Q_i, we have $W(\gamma, P) = (1/2\pi)\Sigma_{i=1}^{r} n_i(\vartheta(Q_i) - \vartheta(P_i))$, which is zero since the boundary of γ is zero. □

6c. Chains on Grids

A *grid* G will be a finite union of lines in the plane, each parallel to x-axis or the y-axis, with at least two horizontal and two vertical lines. These lines cut the plane into a finite number of rectangular regions, some bounded and some unbounded. By a *rectangular* 1-chain for a given grid we shall mean a 1-chain μ of the form $\mu = n_1\sigma_1 + \ . \ . \ . \ + n_r\sigma_r$,

where each σ_j is a straight path along one of the sides of the bounded rectangles, from left to right or from bottom to top.



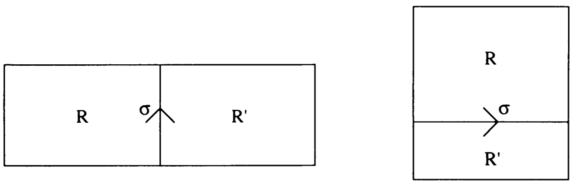
Let $R_1, \ldots, R_r$ be the bounded rectangles for the grid G, numbered in any order, and choose a point P_i in the interior of R_i for each i. The next lemma proves an elementary but important fact: a closed rectangular 1-chain is completely determined by its winding numbers about these points.

Lemma 6.9. *If μ is a closed rectangular 1-chain for G, then*

$$\mu = n_1 \partial R_1 + \ldots + n_r \partial R_r,$$

where $n_i = W(\mu, P_i)$.

Proof. Since the winding number of ∂R_i around P_j is 1 if $i = j$ and 0 otherwise, the winding number of each side of the displayed equation around each P_j is the same. Let $\tau = \mu - \Sigma n_i \partial R_i$ be the difference, which has winding number zero around each P_j. We must show that τ is zero as a 1-chain. Suppose an edge σ occurs in τ with nonzero coefficient m. Suppose that σ is a vertical or horizontal line between two rectangles R and R' of the grid, with R to the left of or above R'. Assume first that R is bounded:



The trick is to consider the closed 1-chain $\tau' = \tau - m \cdot \partial R$. Let P and

P' be interior points in R and R', respectively. Since $W(\partial R, P) = 1$ and $W(\partial R, P') = 0$, we have

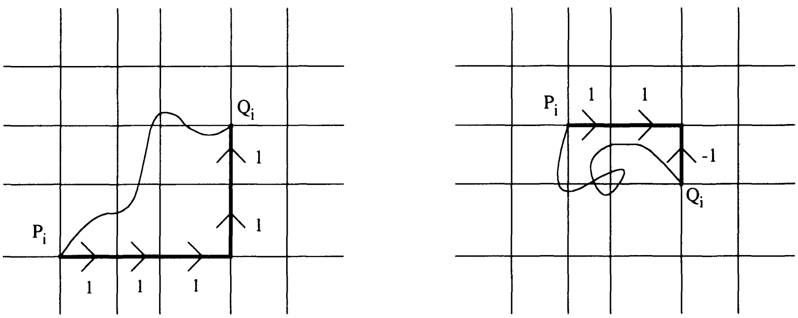
$$W(\tau', P) = W(\tau, P) - m \cdot W(\partial R, P) = -m,$$
$$W(\tau', P') = W(\tau, P') - m \cdot W(\partial R, P') = 0.$$

But the edge σ does not appear in τ', so P and P' belong to the same connected component of the complement of the support of τ', which implies by Proposition 6.8 that $W(\tau', P) = W(\tau', P')$, a contradiction. If R is unbounded, and R' is bounded, the argument is similar, using $\tau' = \tau + m \cdot \partial R'$. □

Next we need an approximation lemma that will assure that, for the purposes of winding numbers and integration, all paths and 1-chains can be replaced by rectangular 1-chains.

Lemma 6.10. *Let γ be any 1-chain in an open set U. Then there is a grid G, and a rectangular 1-chain μ for G, with the support of μ contained in U, such that $\gamma - \mu$ is a 1-boundary in U. If γ is closed, then μ is also closed.*

Proof. By Lemma 6.4(b) we can subdivide any of the paths that occur in γ, and the difference between γ and the 1-chain with subdivided paths will be a boundary. Using the Lebesgue lemma as usual to subdivide, we may therefore assume that each path γ_i occurring in γ maps $[0, 1]$ into some open rectangle U_i contained in U. Let P_i and Q_i be the starting and ending points of γ_i. Take any grid G that has a vertical and horizontal line passing through each P_i and each Q_i. Let μ_i be a rectangular 1-chain for G that goes from P_i to Q_i, involving only edges on the closed rectangle with corners at P_i and Q_i; in particular, the boundary of μ_i is $Q_i - P_i$. (Note that μ_i is in U_i, so μ_i is a chain in U.)



It suffices to verify that $\gamma_i - \mu_i$ is a boundary in U_i, for if $\gamma = \Sigma n_i \gamma_i$, then $\mu = \Sigma n_i \mu_i$ will be the required rectangular 1-chain, with $\gamma - \mu = \Sigma n_i(\gamma_i - \mu_i)$ a boundary. But since $\gamma_i - \mu_i$ is a closed 1-chain on a starshaped open set U_i, this follows from Exercise 6.5. □

We are now ready to prove the main goal of this chapter: that the geometric condition for 1-cycles to be homologous in an open set is equivalent to the numerical condition of having the same winding number around all points outside the open set.

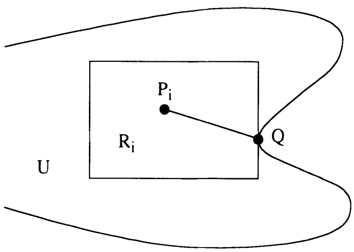
Theorem 6.11. *Suppose γ and δ are closed 1-chains on an open set U in the plane. Then the following are equivalent:*

(1) *γ and δ are homologous, i.e., $\gamma - \delta$ is a boundary in U; and*

(2) *$W(\gamma, P) = W(\delta, P)$ for all points P not in U.*

Proof. We saw in Lemma 6.7 that (1) implies (2). For the converse, by looking at $\tau = \gamma - \delta$, it suffices to show that if τ is a closed 1-chain such that $W(\tau, P) = 0$ for all $P \notin U$, then τ is a boundary. By Lemma 6.10, there is a closed rectangular 1-chain μ for some grid G so that $\tau - \mu$ is a boundary in U. By Lemma 6.9, $\mu = \Sigma n_i \partial R_i$, where $n_i = W(\mu, P_i)$.

We claim that $n_i = 0$ unless the entire closed rectangle R_i is contained in U. For if Q is a point of R_i that is not in U, then $W(\tau, Q) = 0$ by assumption, so $W(\mu, Q) = 0$ by Lemma 6.7. Since the straight line from P_i to Q lies in the complement of the support of μ, it follows from Proposition 6.8 that $W(\mu, P_i) = W(\mu, Q)$, so $n_i = W(\mu, P_i) = 0$.



It follows that $\mu = \Sigma n_i \partial R_i$, with each R_i contained in U, and such a 1-chain is visibly a boundary in U. So $\tau = (\tau - \mu) + \mu$ is also a boundary. □

For example, if $U = \mathbb{R}^2 \setminus \{P\}$ is the complement of a point, two 1-cycles are homologous in U exactly when their winding numbers

around P coincide. In other words, the winding number gives an isomorphism

$$H_1(\mathbb{R}^2 \setminus \{P\}) \xrightarrow{\cong} \mathbb{Z}, \qquad [\gamma] \mapsto W(\gamma, P),$$

where $[\gamma]$ denotes the class of a 1-cycle γ in the homology group.

Exercise 6.12. Suppose $U = \mathbb{R}^2 \setminus \{P_1, \ldots, P_n\}$ is the complement of n points in the plane. Show that the mapping that takes a closed 1-chain γ to $(W(\gamma, P_1), \ldots, W(\gamma, P_n))$ determines an isomorphism of H_1U with the free abelian group $\mathbb{Z}^n$.

Exercise 6.13. State and prove the analogue of Theorem 6.11 when γ and δ are arbitrary 1-chains in U with the same boundary.

6d. Maps and Homology

If $\gamma = n_1\gamma_1 + \ldots + n_r\gamma_r$ is a 1-chain in an open set U, with γ_i paths, and $F\colon U \to U'$ is a continuous mapping from U to another open set U', define $F_*\gamma$ to be the 1-chain in U' defined by

$$F_*\gamma = n_1(F \circ \gamma_1) + \ldots + n_r(F \circ \gamma_r).$$

F also maps 0-chains in U to 0-chains in U': $F_*(\Sigma m_i P_i) = \Sigma m_i F(P_i)$.

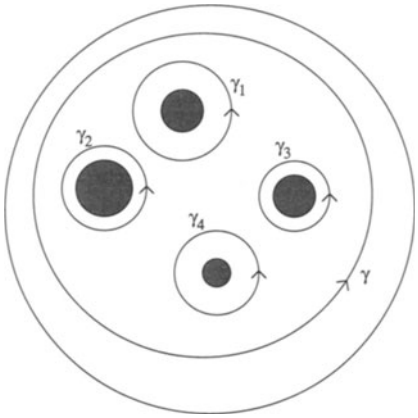
Exercise 6.14. Show that $\gamma \mapsto F_*\gamma$ is a homomorphism from the group of 1-chains in U to the group of 1-chains in U'. If γ is a closed 1-chain, show that $F_*\gamma$ is also closed. Show in fact that $F_*(\partial\gamma) = \partial(F_*\gamma)$ for any 1-chain. Show that if γ is a boundary in U, then $F_*\gamma$ is a boundary in U'.

From this exercise and Theorem 6.11, we deduce the following fact, which is not so obvious from the definition of the winding number:

Proposition 6.15. *If γ and δ are closed 1-chains in U with the same winding number around all points not in U, then $F_*\gamma$ and $F_*\delta$ are closed 1-chains in U' with the same winding number around all points not in U'.* □

For example, take U to be the region inside one disk D and outside a disjoint union of closed disks $A_1, \ldots, A_n$ contained in D. Let γ

be a circular path in U containing the disks A_i, and let γ_i be a circular path around A_i.



Since γ and $\Sigma\gamma_i$ have the same winding numbers around each point not in U, we conclude that for any $F\colon U \to U'$, and any $Q \notin U'$,

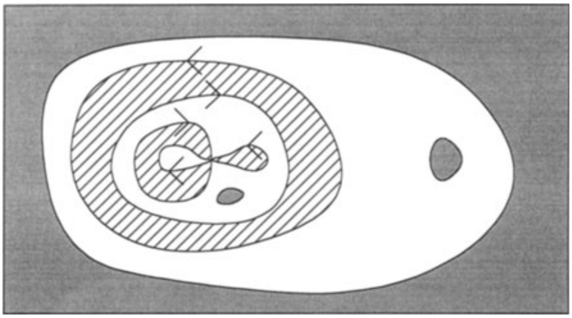
$$W(F_*\gamma, Q) \;=\; W(F\circ\gamma_1, Q) + \;.\;.\;.\; + W(F\circ\gamma_n, Q).$$

Here is an application, which is a substantial generalization of Exercise 4.9.

Corollary 6.16. *Let $F\colon U \to \mathbb{R}^2$ be a continuous mapping, and suppose γ is a closed 1-chain in U such that $W(\gamma, P) = 0$ for all points P not in U. Let Q be a point in $\mathbb{R}^2$, not in $F(Supp(\gamma))$, such that $W(F_*\gamma, Q) \neq 0$. Then there is a point P in U such that $W(\gamma, P) \neq 0$ and $F(P) = Q$.*

Proof. Let $Z = F^{-1}(Q)$, a closed (conceivably empty) subset of U disjoint from Supp(γ). Apply the proposition to the open set $U^\circ = U \setminus Z$, the restriction of F to U°, and the 1-chains γ and $\delta = 0$. If the assertion of the corollary is false, then $W(\gamma, P) = 0$ for all P in Z, so $W(\gamma, P) = 0$ for all P not in U°, so γ is homologous to 0 in U°. But since Q is not in $F(U^\circ)$, the fact that $W(F_*\gamma, Q) \neq 0$ shows that $F_*\gamma$ is not homologous to 0 in $F(U^\circ)$, contradicting Proposition 6.15. □

For example, with U the complement of the gray area, the 1-chain γ that is the sum of the paths shown has a nonzero winding number only around points in the striped area, so any point with $W(F_*\gamma, Q) \neq 0$ would be the image of a point from one of the striped regions.



Problem 6.17. Under the conditions of the corollary, assume that, at each point P of $F^{-1}(Q)$, the local degree of F at P is defined. Show that

$$W(F_*\gamma, Q) = \sum_{P \in F^{-1}(Q)} \deg_P(F) \cdot W(\gamma, P).$$

Problem 6.18. Let E be the closed set obtained from a closed disk D by removing the interiors of k disjoint disks $D_1, \ldots, D_k$ contained in D, for some $k \geq 0$. Let C be the boundary of D, and C_i the boundary of D_i. Suppose $F: E \to \mathbb{R}^2$ is a continuous mapping such that for each point P in any of the boundary circles, the vector from P to $P + F(P)$ is not tangent to that circle. Show that, if $k \neq 1$, there must be a point Q in E with $F(Q) = 0$. What about the case $k = 1$?

6e. The First Homology Group for General Spaces

The general definitions of this chapter make sense without change for any topological space, although, of course, one does not have winding numbers in arbitrary spaces. Any topological space X has abelian groups Z_0X of 0-chains, B_0X of 0-boundaries, with 0th homology group $H_0X = Z_0X/B_0X$, which is canonically isomorphic to the free abelian group on the set of path-connected components of X. Similarly, one has the abelian group C_1X of 1-chains on X, with the subgroups Z_1X of 1-cycles and 1-boundaries B_1X, and the 1st homology group $H_1X = Z_1X/B_1X$. There are no changes in the definitions, other than replacing a "U" by an "X."

Exercise 6.19. Show that a continuous mapping $F: X \rightarrow Y$ determines a homomorphism from Z_1X to Z_1Y taking B_1X to B_1Y.

This determines a homomorphisms of the quotient groups, denoted

$$F_*: H_1X \rightarrow H_1Y.$$

Exercise 6.20. Show that if $F: X \rightarrow Y$ and $G: Y \rightarrow Z$ are continuous, then $(G \circ F)_* = G_* \circ F_*$ as homomorphisms from H_1X to H_1Z. If F is the identity map on X, show that F_* is the identity map on H_1X.

The result of this exercise is expressed by saying that the diagram

$$\begin{array}{ccc} H_1X & \xrightarrow{\;F_*\;} & H_1Y \\ & {\scriptstyle (G\circ F)_*}\searrow \quad \swarrow{\scriptstyle G_*} & \\ & H_1Z & \end{array}$$

commutes: starting with an element in the upper left group H_1X, mapping it to H_1Z by either route gives the same answer.

For example, if $Z = X$, and F and G are homeomorphisms that are inverses to each other, so that $G \circ F$ is the identity map of X, then $(G \circ F)_*$ is the identity map of H_1X, so the composite

$$H_1X \xrightarrow{F_*} H_1Y \xrightarrow{G_*} H_1X$$

is the identity map on H_1X. Similarly $F \circ G$ is the identity map on Y, so $G_* \circ F_*$ is the identity map on H_1Y. It follows that F_* and G_* are inverse isomorphisms between H_1X and H_1Y. In particular,

Proposition 6.21. *If X and Y are homeomorphic*, then H_1X *and* H_1Y *are isomorphic abelian groups.*

Similarly, if Y is contained in X, and $r: X \rightarrow Y$ is a continuous retract, the identity map on H_1Y must factor into a composite of homomorphisms $H_1Y \rightarrow H_1X \rightarrow H_1Y$. For example, if $H_1X = 0$, and $H_1Y \neq 0$, this shows that there can be no such retract.

Exercise 6.22. Compute H_1X for X a circle and for X a disk, and show again that a circle is not a retract of the disk it bounds.

Two continuous maps F and G from X to Y are *homotopic* if there

is a continuous mapping $H: X \times [0, 1] \to Y$ such that $F(x) = H(x, 0)$ and $G(x) = H(x, 1)$ for all x in X.

Proposition 6.23. *If F and G are homotopic maps from X to Y, then $F_* = G_*$.*

Proof. If $\gamma = \Sigma n_i \gamma_i$ is a 1-cycle on X, set $\Gamma_i(t, s) = H(\gamma_i(t), s)$. Then $F_*\gamma - G_*\gamma = \Sigma n_i \partial \Gamma_i$, the other terms canceling each other since γ is a cycle. □

A subspace Y of a space X is called a *deformation retract* if there is a continuous retract $r: X \to Y$ such that the identity map from X to X is homotopic to the map $i \circ r$, where i is the inclusion of Y in X. A space X is called *contractible* if it contains a point that is a deformation retract of X.

Exercise 6.24. (a) Show that the circle $S^1 \subset \mathbb{R}^2 \setminus \{0\}$ is a deformation retract. (b) Give an example of a retract that is not a deformation retract. (c) Show that any two maps from any space to a contractible space are homotopic. In particular, every point in a contractible space is a deformation retract of the space.

Exercise 6.25. Show that if Y is a deformation retract of X, then the map from H_1Y to H_1X determined by the inclusion of Y in X is an isomorphism. Show that $H_1(X) = 0$ if X is contractible.

Exercise 6.26. Show that $F: X \to Y$ determines a homomorphism from Z_0X to Z_0Y taking B_0X to B_0Y, and so a homomorphism, also denoted F_*, from H_0X to H_0Y. Verify the analogues of the assertions in Exercises 6.19 and 6.20, and Propositions 6.21 and 6.23.

If X is a subspace of the plane that is not open, one does have a notion of the winding number around points not in X, but the situation is more complicated, as the following problems indicate.

Problem 6.27 (For those who know the Tietze extension theorem). Suppose X and X' are closed subsets of the plane, and $F: X \to X'$ is a continuous mapping. Suppose γ and δ are two closed 1-chains on X such that $W(\gamma, P) = W(\delta, P)$ for all P not in X. Show that $W(F_*\gamma, P') = W(F_*\delta, P')$ for all P' not in X'.

Problem 6.28. Let X be a closed set in the plane. Show that if γ and δ are homologous 1-cycles on X, then $W(\gamma, P) = W(\delta, P)$ for all P not in X. Is the converse true?

PART IV
VECTOR FIELDS

A mapping from an open set in the plane to the plane can be regarded as a vector field, and winding numbers can be used to define the index of a vector field at a singularity. The ideas of Chapter 6 can be used to relate sums of indices to winding numbers around regions. This is applied to show that vector fields on a sphere must have singularities: one cannot comb the hair on a billiard ball. The same ideas are used in the next chapter to study more interesting surfaces. These chapters are inserted here to indicate some other interesting things one can do with winding numbers; a reader in a hurry to move on can skip to Part V or VI.

Chapter 8 sketches how some of the ideas we have studied in the plane and on the sphere can be studied on more general surfaces. It gives us a first chance to study some spaces other than plane regions and spheres. In particular, we see how the "global" topology of the surface puts restrictions on the "local" data of indices of a vector field. We use this to discuss the Euler characteristic of a surface. Some of the arguments in this section will depend on geometric constructions that will only be sketched, usually by drawing pictures showing how to deform one surface into another. Later in the chapter we discuss briefly what it would take to make these arguments rigorous, and later in the book we take up the study of surfaces more systematically.

CHAPTER 7

Indices of Vector Fields

7a. Vector Fields in the Plane

We want to look at continuous vector fields on an open set U in the plane, but allowing them to have a finite number of *singularities*. A singularity will be a point at which either the vector field is not defined, or a point where it is defined and is zero. A *vector field on U with singularities in Z* will therefore be a continuous mapping

$$V\colon\ U \setminus Z \to \mathbb{R}^2 \setminus \{0\},$$

where Z is a finite set in U. For vector fields arising from flow of a fluid, the singularities may arise from "sources," where fluid is entering the system, or "sinks," where it is leaving, or some other discontinuity.

Given such a vector field V, to each point P in U one can define an integer called the *index of V at P*, denoted $\text{Index}_P V$. To do this, take a disk D_r of some radius r about P that does not meet Z at any point except (perhaps) P. Let C_r be the boundary of this disk. The restriction $V|_{C_r}$ of V to C_r is a mapping from this circle to $\mathbb{R}^2 \setminus \{0\}$, so it has a winding number (by §3d). Define the index to be this winding number:

$$\text{Index}_P V \;=\; W(V|_{C_r}, 0)\,.$$

Lemma 7.1. (*a*) *This definition is independent of choice of r.*

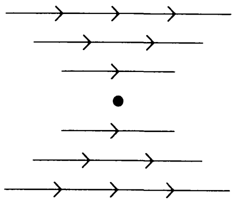
(*b*) *If P is not in Z, then* $\text{Index}_P V = 0$.

Proof. The proof of (a) is the same as that of Lemma 3.28, and (b) follows from Proposition 3.20. □

From the definition, the index at P depends only on the restriction of V to an arbitrarily small neighborhood of P. Here are some examples of vector fields with singularities at the origin, with the corresponding indices; the calculations are left as exercises. Instead of drawing vectors at many points, it is more useful to draw some flow lines, i.e., curves that are tangent to the vector field at each point.

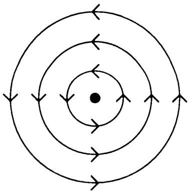
Vector field V(x, y) — *Index at* 0

(1) $(x^2+y^2, 0)$



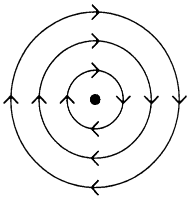
0

(2) $(-y, x)$



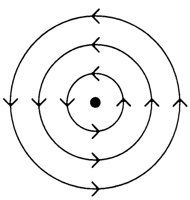
1

(3) $(y, -x)$



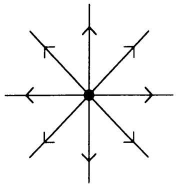
1

(4) $\left(\dfrac{-y}{x^2+y^2}, \dfrac{x}{x^2+y^2}\right)$



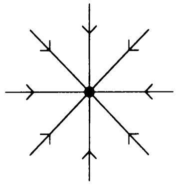
1

(5) (x, y)



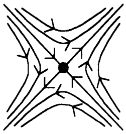
1

(6) $(-x, -y)$



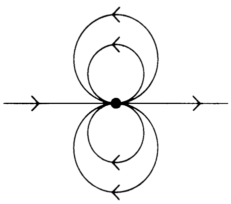
1

(7) (y, x)



-1

(8) $(x^2 - y^2, 2xy)$



2

Note that (2) and (3) are opposite as vector fields, as are (5) and (6), but have the same index. Similarly, (4) is a positive multiple of (2). These are special cases of Exercise 7.4, which shows that the magnitude and sign of the vectors does not affect the index (which explains why the flow lines, without even their sense of direction, determine the indices).

Exercise 7.2. Construct, for each integer n, a vector field with a singularity at the origin with index n.

Exercise 7.3. Let V_0 and V_1 be continuous vector fields in a punctured neighborhood U of P, and suppose there is a continuous mapping $H: U \times [0, 1] \to \mathbb{R}^2$ such that

$$H(Q \times 0) = V_0(Q), \qquad H(Q \times 1) = V_1(Q) \qquad \text{for all } Q \text{ in } U;$$
$$H(Q \times t) \neq 0 \qquad \text{for all } Q \text{ and all } 0 \leq t \leq 1.$$

Show that V_0 and V_1 have the same index at P.

Exercise 7.4. Show that if ρ is any continuous function defined in a punctured neighborhood of P that is always positive or always negative in this neighborhood, then the index at P of $\rho \cdot V$ is the same as the index at P of V.

The following proposition relates the behavior of a vector field along the boundary of a region to the indices of the vector field at singular points inside:

Proposition 7.5. *Let V be a vector field with singularities in U. Suppose γ is a closed 1-chain in U whose support does not meet the singular set Z of V, such that $W(\gamma, P) = 0$ for all P not in U. Then*

$$W(V_*\gamma, 0) = \sum_{P \in Z} W(\gamma, P) \cdot \mathrm{Index}_P V.$$

Proof. Let $Z = \{P_1, \ldots, P_r\}$, and let $D_1, \ldots, D_r$ be disjoint closed disks centered at the points $P_1, \ldots, P_r$, all contained in U. Let γ_i be the standard counterclockwise path around the boundary of D_i, and let $n_i = W(\gamma, P_i)$. Then γ and $n_1\gamma_1 + \ldots + n_r\gamma_r$ have the same winding number around every point outside $U \setminus Z$. It follows from Proposition 6.15 that

$$W(V_*\gamma, 0) = n_1 W(V \circ \gamma_1, 0) + \ldots + n_r W(V \circ \gamma_r, 0),$$

and this is the assertion to be proved. □

Corollary 7.6. *If $W(V_*\gamma, 0) \neq 0$, then V must have at least one nonvanishing index at a point P with $W(\gamma, P) \neq 0$.* □

The simplest case of the proposition is:

Corollary 7.7. *If U contains a closed disk D, and V has no singularities on the boundary circle C of D. Then*

$$W(V|_C, 0) = \sum_{P \in D} \mathrm{Index}_P V.$$

If $W(V|_C, 0) \neq 0$, V must have a singularity with nonvanishing index inside D. □

Problem 7.8. Generalize this discussion to allow the singularity set Z to be infinite but *discrete*, i.e., Z is a closed subset of U such that each P in Z has a neighborhood U_P in U such that $U_P \cap Z = \{P\}$. Show that the sum in Proposition 7.5 is automatically finite.

Problem 7.9. Let f be a $\mathscr{C}^\infty$ function defined in U, defining a gradient vector field $V = \text{grad}(f)$ on U. A point P is a *critical point* for f if the gradient vanishes at P, and P is called *nondegenerate* if the "Hessian" at P,

$$\frac{\partial^2 f}{\partial x^2}(P) \cdot \frac{\partial^2 f}{\partial y^2}(P) - \left(\frac{\partial^2 f}{\partial x \, \partial y}(P)\right)^2$$

is not zero. (a) Show that, if P is a nondegenerate critical point, then

$$\text{Index}_P(V) = \begin{cases} 1 & \text{if } f \text{ has a local maximum or minimum at } P, \\ -1 & \text{otherwise (when } P \text{ is a saddle point).} \end{cases}$$

(b) Suppose D is a disk in U with boundary C, and f has only nondegenerate critical points in D, with none on the boundary, and f is constant on C. Show that the number of local maxima plus the number of local minima is one more than the number of saddle points. "On a circular island, the number of peaks plus the number of valleys is one more than the number of passes."

7b. Changing Coordinates

Later we want to define the index of a vector field on a surface other than an open set in the plane. To do this, the essential point is to know that a change of coordinates does not change the index. This result, which is intuitively obvious from pictures of vector fields, takes some care to state and a little work to prove. To state it, suppose $\varphi\colon U \to U'$ is a diffeomorphism from one open set in the plane onto another; that is, φ is $\mathscr{C}^\infty$, one-to-one, and onto, and the inverse map $\varphi^{-1}\colon U' \to U$ is also a $\mathscr{C}^\infty$ mapping. At any point P in U, we have the Jacobian matrix

$$J_{\varphi,P} = \begin{bmatrix} \dfrac{\partial u}{\partial x}(P) & \dfrac{\partial u}{\partial y}(P) \\[2ex] \dfrac{\partial v}{\partial x}(P) & \dfrac{\partial v}{\partial y}(P) \end{bmatrix},$$

where $\varphi(x, y) = (u(x, y), v(x, y))$ in coordinates. This gives a linear mapping from vectors in $\mathbb{R}^2$ to vectors in $\mathbb{R}^2$ (see Appendix C). If V is a continuous vector field in U, define the vector field $\varphi_* V$ in U'

by the formula

$$(\varphi_* V)(P') = J_{\varphi,P}(V(P)),$$

where P is the point in U mapped to P' by φ, i.e., $P = \varphi^{-1}(P')$. If V has singularities in the set Z, V' will have singularities in $\varphi(Z)$.

Lemma 7.10. *With V and $\varphi_* V$ as above, then, for any P in U,*

$$\text{Index}_{\varphi(P)}(\varphi_* V) = \text{Index}_P V.$$

The proof of this lemma is given in Appendix D. We will also want to compare the indices of two different vector fields on a surface. The following lemma will be used to reduce to the case where they agree in a neighborhood of some point. We say that two vector fields V and W *agree* on a set A if $V(P) = W(P)$ for all P in A. The proof is also in Appendix D.

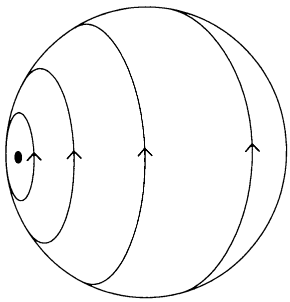
Lemma 7.11. *Suppose V and W are continuous vector fields with no singularities on an open neighborhood U of a point P. Let $D \subset U$ be a closed disk centered at P. Then there is a vector field $\tilde{V}$ with no singularities on U such that: (i) $\tilde{V}$ and V agree on $U \setminus D$; and (ii) $\tilde{V}$ and W agree on some neighborhood of P.*

7c. Vector Fields on a Sphere

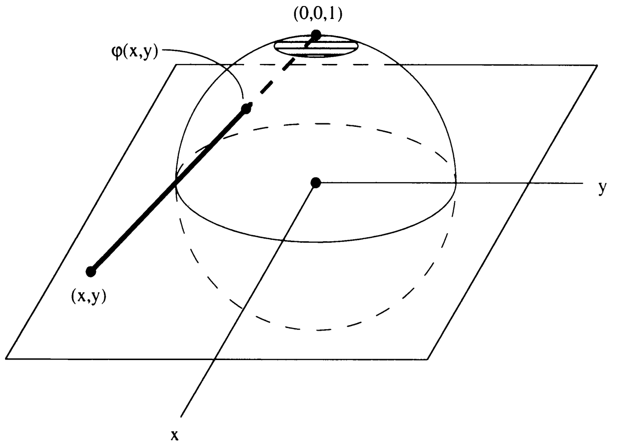
A vector field V on a sphere S assigns to each point P in S a vector $V(P)$ in the tangent space $T_P S$ to S at P, the mapping from P to $V(P)$ being continuous. If $S = S^2$ is the standard sphere, and $P = (x, y, z)$, the tangent space is

$$T_P S = \{(a, b, c): (a, b, c) \cdot (x, y, z) = ax + by + cz = 0\}.$$

As before, we may allow a finite set Z of singularities. In fact, one of our goals is to show that any vector field on sphere *must* have singularities.



We want to flatten out the sphere, say by stereographic projection from a point on the sphere, so that the vector field determines a corresponding vector field on the plane, and we will use what we know about vector fields on the plane.



We need the following, whose proof is left as an exercise:

Lemma 7.12. *The inverse φ of this polar coordinate mapping takes (x, y) in the plane to*

$$\varphi(x, y) = \left(\frac{2x}{x^2 + y^2 + 1}, \frac{2y}{x^2 + y^2 + 1}, \frac{x^2 + y^2 - 1}{x^2 + y^2 + 1}\right) \quad \text{in} \quad S^2.$$

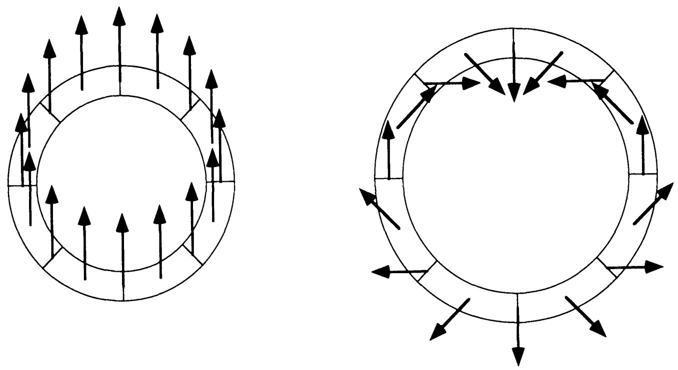
The Jacobian matrix $J_{\varphi,P}$ of φ at $P = (x, y)$ maps $\mathbb{R}^2$ one-to-one onto the tangent space to S^2 at the point $\varphi(P)$. If V is a vector field on S^2, and φ^*V is the vector field on $\mathbb{R}^2$ defined by the equation

$$J_{\varphi,P}((\varphi^*V)(P)) = V(\varphi(P)),$$

then φ^*V is continuous at P if V is continuous at $\varphi(P)$.

Now suppose V is a continuous vector field with no singularities on S^2. Then φ^*V is a vector field on $\mathbb{R}^2$ with no singularities. Let C_r be a large circle centered at the origin in the plane. We know from Corollary 7.7 that the winding number of φ^*V around C_r must be zero. To get a contradiction, we must use the fact that V is continuous and nonzero at the north pole. For r large, C_r can be thought of as a small circle around the north pole. The winding number of φ^*V around such a circle is not zero, even though it comes from a vector field

that is not zero in the disk near the north pole. This can be seen by unraveling what happens to a vector field on the sphere when the sphere is flattened out. Think of the vectors as attached to a small band, which is turned over:



The winding number of this vector field around this circle is 2, which shows that no such vector field can exist.

Problem 7.13. Give an analytic proof of the fact that this winding number is 2.

This shows more than the fact that a vector field on a sphere must have singularities. If V is a vector field on S^2 with singularities in a finite set Z that does not include the north pole, we can define the index $\text{Index}_P V$ at a point P of Z to be the index of $\varphi^* V$ at the corresponding point $\varphi^{-1}(P)$. Then the proof shows that, for any such vector field V on S^2,

$$\sum_{P \in Z} \text{Index}_P V = 2.$$

In order to include the north pole in these considerations, we also consider a stereographic projection from the south pole. To make the orientations match in our two charts, we first reflect in the x-axis. So we define $\psi\colon \mathbb{R}^2 \to S^2$ by

$$\psi(x, y) = \left(\frac{2x}{x^2+y^2+1}, \frac{-2y}{x^2+y^2+1}, \frac{-x^2-y^2+1}{x^2+y^2+1}\right).$$

Exercise 7.14. Show that the composite $\varphi^{-1} \circ \psi$ takes $z = x + iy$ to $1/z$.

Now for any vector field V with a finite number of singularities on S^2, and any point P, we can define $\text{Index}_P V$ as either the index of $\varphi^* V$ or of $\psi^* V$ at the corresponding point. It follows from Lemma 7.10 that these indices agree, if both are defined. In fact, we could use a stereographic projection from any point; all the coordinate transformations as above are $\mathscr{C}^\infty$.

Proposition 7.15. *For any vector field with singularities V on S^2,*

$$\sum_{P \in Z} \text{Index}_P V = 2.$$

Proof. Instead of arguing as we did above, we can argue in two steps:

Step 1. There is a vector field V on S^2 with $\Sigma_{P \in Z} \text{Index}_P(V) = 2$. For example, if W is the vector field on $\mathbb{R}^2$ given in (8), i.e., $W(x, y) = (x^2 - y^2, 2xy)$, then $V = \varphi_* W$ is a vector field on the complement of the north pole with one singularity of index 2 at the south pole. A short calculation shows that V extends continuously to the north pole, with value there the vector $(-2, 0, 0)$.

Step 2. We show that the sum of the indices of any two such vector fields V and W on S^2 is the same. Let P be a point where neither has a singularity. By Lemma 7.11, replacing V by another vector field $\tilde{V}$ with the same indices as V, we can assume that V and W agree in some neighborhood of P. Then using stereographic projection from P, one has two vector fields on the plane that agree outside some large disk that contains all the singularities of either vector field. Taking a larger circle C_r, their winding numbers around C_r will be the same, and an application of Corollary 7.7 shows that the sum of their indices is the same. □

Exercise 7.16. Give an alternative proof of Step 1 by finding a vector field on S^2 with two singular points, each with index 1.

Problem 7.17. (a) If $f\colon S^2 \to \mathbb{R}^3$ is a continuous mapping, show that there is some point P in S^2 and some real number λ so that $f(P) = \lambda P$. (b) If $f\colon S^2 \to S^2$ is a continuous mapping, show that there is some point P in S^2 such that $f(P) = P$ or $f(P) = -P$.

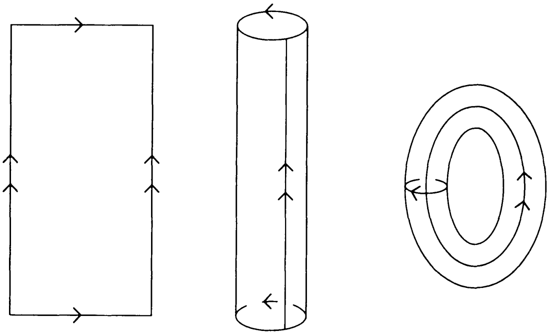
Problem 7.18. Give a mathematical formulation and proof of the statement: "On a spherical planet, the number of peaks plus the number of valleys is two more than the number of passes."

CHAPTER 8

Vector Fields on Surfaces

8a. Vector Fields on a Torus and Other Surfaces

Let us look next at a surface X which is a *torus*, i.e., the surface of a doughnut. This can be realized concretely in several ways: as a surface of revolution, by an explicit equation in 3-space, as a Cartesian product $S^1 \times S^1$, or by taking a square or rectangle and identifying opposite edges:



It is clear by any of these descriptions that there are vector fields on X that have no singularities at all. In order to state the analogue of Proposition 7.15 for a torus, we need to define the index of a vector field at a point on X. One way to do this is to realize X as a quotient

space of the plane $\mathbb{R}^2$, identifying two points if their difference is in the lattice $\mathbb{Z}^2$. This amounts to identifying the opposite sides of the unit square $[0,1] \times [0,1]$. We have a mapping

$$p\colon \mathbb{R}^2 \to X = S^1 \times S^1,$$

and giving a vector field V on X is the same as giving a vector field $\tilde{V}$ on $\mathbb{R}^2$ that is unchanged by translation by any vector in $\mathbb{Z}^2$. We can define the index $\text{Index}_P V$ to be the index of the corresponding vector field $\tilde{V}$ at any point of $\mathbb{R}^2$ that maps to P. As before, we allow a finite set Z of singularities, and require V to be continuous outside Z. Then we have the analogous proposition:

Proposition 8.1. *For any vector field with singularities V on a torus X,*

$$\sum_{P \in Z} \text{Index}_P V = 0.$$

Proof. Take a square $R = [a, a+1] \times [b, b+1]$ so that the image in X of the boundary ∂R does not hit the singularity set Z. Look at the corresponding vector field $\tilde{V}$ on a neighborhood of R. By Proposition 7.5 the winding number of $\tilde{V}$ around ∂R is the sum of the indices of $\tilde{V}$ inside R. This winding number is zero, since the vector field is the same on opposite sides of the square. And the indices inside are the indices of V on X. □

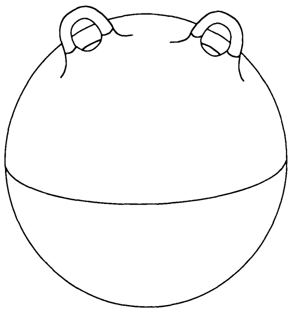
For example, realizing X as a surface of revolution in space as in the above picture, the projection of a vertical vector $(0,0,1)$ onto the tangent space at each point gives a vector field with singularities at the four points with horizontal tangents. The indices at the points at the top and bottom (where the height has local maximum and minimum) are $+1$, while those at the two saddle points are -1.

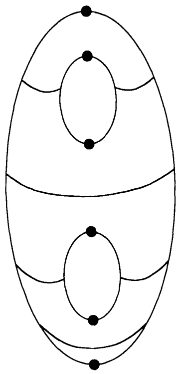
Exercise 8.2. Show that the following formula defines a diffeomorphism of $\mathbb{R}^2/\mathbb{Z}^2 = S^1 \times S^1$ with a surface of revolution in $\mathbb{R}^3$. Take $0 < r < R$ and define the map by

$$(x, y) \mapsto (R + r\cos(2\pi y)) \cdot (0, \cos(2\pi x), \sin(2\pi x)) + r\sin(2\pi y) \cdot (1, 0, 0).$$

Find the vector field on the plane corresponding to the above vector field, and verify that the indices are 1, -1, -1, and 1.

We want to generalize what we have just seen from the sphere and torus to other surfaces, in particular to the surface of a doughnut with g holes, or a "sphere with g handles":



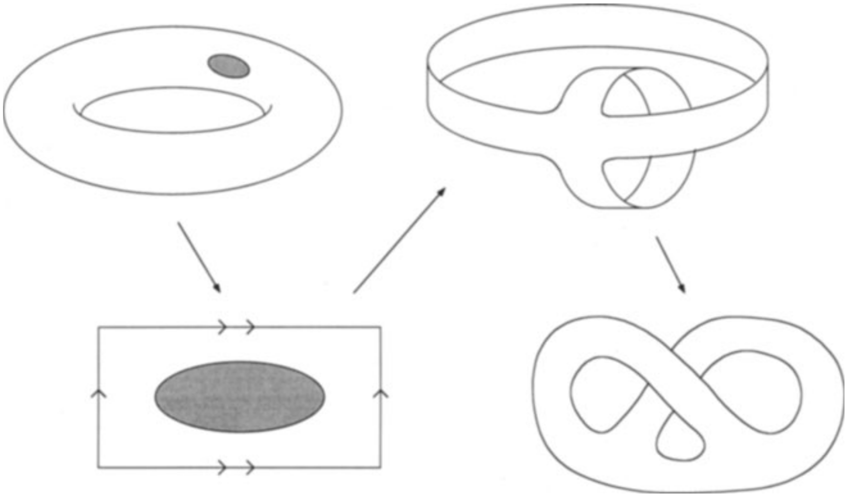


We will argue geometrically and rather loosely for now, and postpone until later a discussion of what needs to be done to make the arguments precise and rigorous. From the second picture, taking V to be the vector field with $V(P)$ the projection on the tangent space T_PX of a vertical vector $(0, 0, 1)$ as before, we see that there are two points with index 1 (at the top and bottom), and $2g$ points with index -1 (at the saddles). This gives one vector field the sum of whose indices is $2 - 2g$. The claim is that this is always the case.

Theorem 8.3 (Poincaré–Hopf). *Let X be a sphere with g handles. For any vector field V with singularities on X, the sum of the indices of V at the singular points is $2 - 2g$.*

Proof. Having seen one such vector field, it is enough to show that any two vector fields have the same sum of indices. By Lemma 7.11, we can take a disk in X where both have no singularities, and modify one so that they agree on such a disk. The idea of the proof is to mimic the proof for a sphere: take a circle C in such a disk, and punch out a smaller disk D (inside the circle) from the surface, and spread the complement $X \setminus D$ out on the plane. The two vector fields will then have the same winding number around the curve C, so they will have the same sum of indices by Proposition 7.5.

For $g > 0$, however, this complement is not diffeomorphic (or homeomorphic) to a plane domain. However, it can be realized with a mapping $\varphi: X \setminus D \to \mathbb{R}^2$ that is a local diffeomorphism, i.e., every point P in $X \setminus D$ has a neighborhood that is mapped diffeomorphically (with a diffeomorphic inverse) onto its image. To visualize this, look first at the torus. The complement of a disk is formed of two bands joined together. The mapping φ from $X \setminus D$ can be visualized by picturing the bands over the plane.

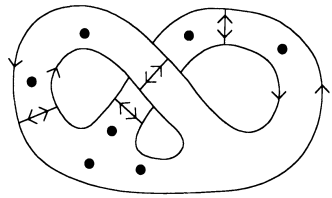


The image of the circle C goes around near the boundary. We can define the winding number $W(V|_C, 0)$ by cutting C up into pieces, each of which is mapped one-to-one by φ into the plane, and using the usual definition on these pieces. We can also define the index of V at a point P of $X \setminus D$ by using the local diffeomorphism φ_* to identify V with a vector field $\varphi_* V$ near $\varphi(P)$, and defining $\text{Index}_P V$ to be the index of $\varphi_* V$ at $\varphi(P)$. To finish the proof, it is enough to show that

$$\sum_{P \in X \setminus D} \text{Index}_P V = W(V|_C, 0).$$

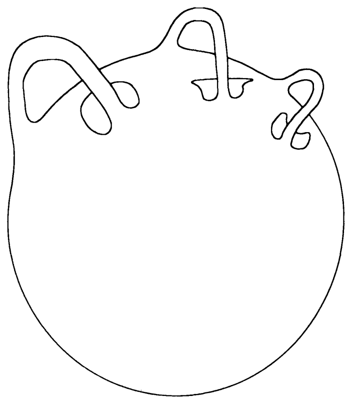
To see this, one can add some crosscuts, being careful not to go through

any singularities, and apply Proposition 7.5 to the restriction of V to each of the pieces.

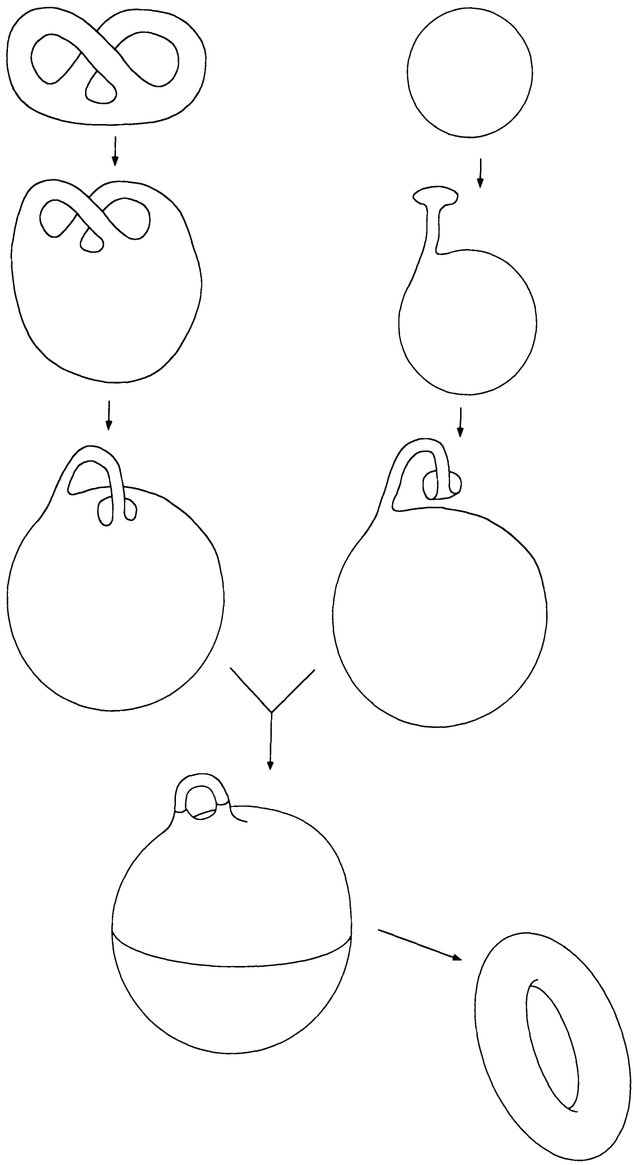


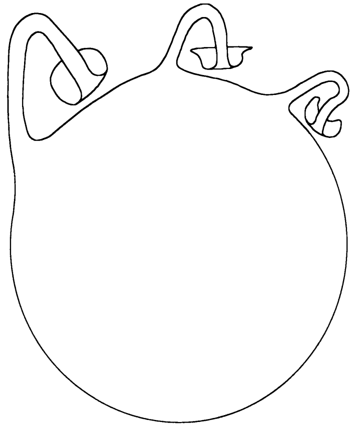
As usual, the added pieces get counted twice with opposite signs, so they cancel, and one is left with the displayed equation.

The same works for any genus g, although the visualization is a little harder. We claim that, when a closed disk D is removed from X, the complement can be realized over the plane as indicated (for $g = 3$):



Once we have this, the same argument completes the proof of the theorem. To realize $X \setminus D$ this way, an essential point is that, as a larger and larger disk is removed from X, the complements are all diffeomorphic. To aid in visualization, the situation for $g = 1$ is redone this way on a separate page. The picture after that shows a disk with "fingers," which can be sewn to the back of the above figure, giving a sphere with g handles.





Exercise 8.4. Describe a vector field on X that has exactly $2g-2$ singular points, each with index -1.

It is time to discuss what it might take to make this sort of argument rigorous. First, of course, one should give precise definitions of all the objects involved: surfaces, a sphere with g handles, a vector field on a surface, and the index of a vector field at a point. This not being a course on manifolds, we will not go through all the details (but see Appendix D for some of them). A key point, however, is that a surface is covered by the images of coordinate charts $\varphi_\alpha\colon U_\alpha \to X$, that are homeomorphisms from open sets U_α in the plane to open sets $\varphi_\alpha(U_\alpha)$ in X, and the change of coordinate mappings $\varphi_\beta^{-1} \circ \varphi_\alpha$, defined from part of U_α to part of U_β, should be $\mathscr{C}^\infty$. For the sphere, for example, stereographic projections from the two poles gave coordinate charts φ and ψ, and for the torus, the mapping $p\colon \mathbb{R}^2 \to X = S^1 \times S^1$, restricted to small open sets in the plane, gives charts on X.

A vector field V on X then determines a vector field V_α on U_α, and these are related by the condition that $(\varphi_\beta^{-1} \circ \varphi_\alpha)_*(V_\alpha) = V_\beta$ on the open subsets where both are defined. In fact, a vector field on X can be defined as a collection of such vector fields V_α, related by these compatibilities under changes of coordinates. The index of V at a point P in X can be defined as the index of V_α at the point P_α, if $\varphi_\alpha(P_\alpha) = P$. The key Lemma 7.10 implies that this is independent of choice of chart near P. Note that the surface X is assumed to have a differentiable structure, but that the vector field is only assumed to be continuous in the complement of a finite set.

The description of surfaces via cutting and pasting, which we have indicated in pictures, could be done explicitly in coordinates, as a

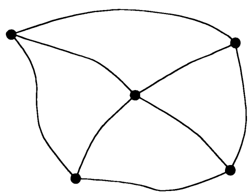
large and not very pleasant exercise. For example, one can verify that removing a slightly larger disk from a surface gives diffeomorphic complements. The tools for doing this sort of argument properly are developed systematically in the subject of differential topology, see Milnor (1965), Wallace (1968), and Guillemin and Pollack (1974).

Exercise 8.5. What can you say about the number of peaks, valleys, and passes on a planet shaped like a sphere with g handles?

Exercise 8.6. Does a sphere with g handles have the fixed point property?

8b. The Euler Characteristic

Suppose X is a compact surface, and we have a *triangulation* of X. This means that X is cut into pieces homeomorphic to triangles, fitting together along the edges. We will have a certain number v of vertices (points), a number e of edges (homeomorphic to closed intervals), and a number f of faces (homeomorphic to closed triangles). These homeomorphisms are assumed to take the ends of intervals to two distinct vertices, and the three boundary pieces of a triangle onto three distinct edges.



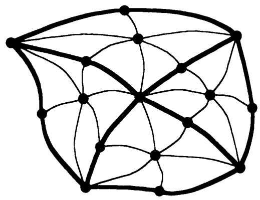
As you may have discovered in an exercise in the Preface, the number $v - e + f$ is independent of the triangulation:

Proposition 8.7. *For any triangulation of a sphere X with g handles,*

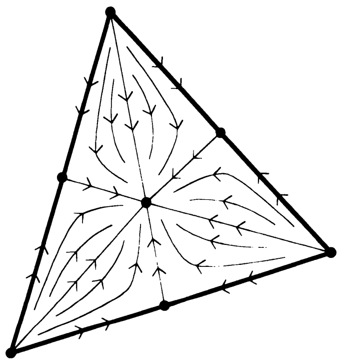
$$v - e + f = 2 - 2g.$$

Proof. The idea is to construct a vector field V on X with one singularity of index 1 for each vertex, one of index -1 for each edge, and one of index 1 for each face. Then the proposition follows from the Poincaré–Hopf theorem. To do this, do a "barycentric subdivi-

sion": put a new vertex in each edge, and one in each face, and connect them as shown.



Then construct a vector field on X, so that in each triangle it looks like:



Construct it first along the edges: on the old edges pointing from the old vertices to the new ones added in the middle of the edges, and on the new edges pointing toward the new vertices in the faces; then fill in over the new triangles to make it continuous. If this is done, one has a vector field V whose singularities are at the vertices, and whose indices are: $+1$ if P is an old vertex; -1 if P is a new vertex along an edge; and $+1$ if P is a new vertex in a face. There are v of the first, e of the second, and f of the third, so by the Poincaré–Hopf theorem, $v\cdot(+1)+e\cdot(-1)+f\cdot(+1)=2-2g$. □

The number $v-e+f$, for any triangulation, which is the same as the sum of the indices of any vector field, is called the *Euler characteristic* of the surface.

Exercise 8.8. Construct a triangulation on the sphere with g handles, and verify the proposition for this triangulation.

Problem 8.9. (a) Show that, for any triangulation of a sphere with g handles:

(i) $2e = 3f$;
(ii) $e \leq {}^1\!/_2 v \cdot (v-1)$; and
(iii) $v \geq {}^1\!/_2(7 + \sqrt{49 - 24(2-2g)})$.

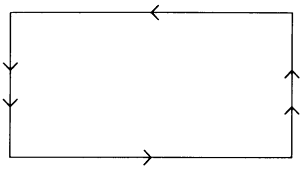
(b) Find lower bounds for v, e, and f for $g = 0$ and $g = 1$, and construct triangulations achieving these lower bounds.

If N is the largest integer less than or equal to the number ${}^1\!/_2(7 + \sqrt{49 - 24(2-2g)})$, it is a fact that any map on X can be colored with N colors, and N is the smallest number for which this is true. See Rademacher and Toeplitz (1957) and Coxeter (1989). Surprisingly, this is *much* easier for $g > 0$ than for $g = 0$.

Exercise 8.10. Generalize Proposition 8.7 to allow arbitrary convex polygons for faces in place of triangles.

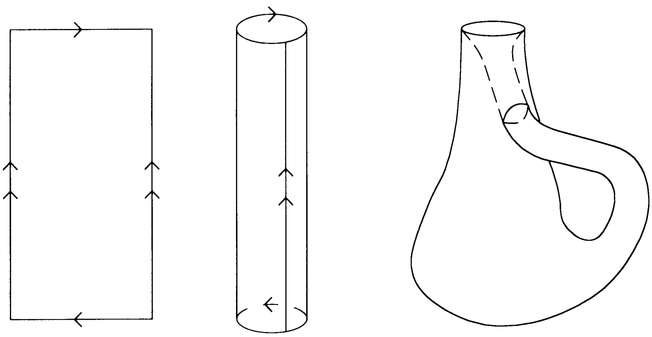
Problem 8.11. Suppose a sphere with g handles is decomposed as in the preceding exercise, but with each polygon having the same number p of edges, and assume that each vertex lies on the same number q of edges. (a) Show that $1/p + 1/q = {}^1\!/_2 + (1-g)/e$. (b) When $g = 0$, find all positive integers p, q, and e that satisfy this equation, and show that all possibilities are realized by the boundaries of the five platonic (regular) solids.

A sphere with g handles can be *oriented*, i.e., one can coherently (continuously) define a notion of counterclockwise (or "which way is up") in the neighborhood of any point. (See Appendix D for a precise definition.) It is a fact that the only compact surfaces that can be oriented are diffeomorphic to spheres with g handles. At least with the assumption that the surface can be triangulated we will prove this in Chapter 17. There are also compact surfaces that cannot be oriented. One example is the *projective plane* $\mathbb{RP}^2$, which can be realized as a quotient space of the sphere S^2, by identifying each point with its antipodal point. The projection from S^2 to $\mathbb{RP}^2$ is two-to-one, but a local diffeomorphism. One can also get $\mathbb{RP}^2$ by taking the upper hemisphere, and identifying opposite points on the boundary equator, which is the same as identifying opposite points on the boundary of a disk. Hence the projective plane can also be realized by identifying the opposite sides of a rectangle as shown:



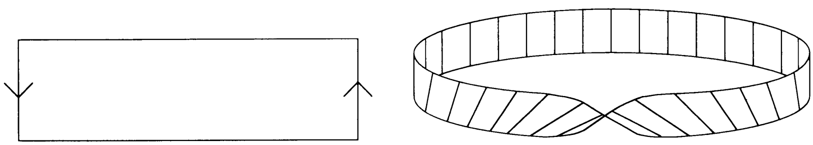
Exercise 8.12. Construct a vector field on $\mathbb{RP}^2$ that has one singular point with index 1. Show that the sum of the indices of any vector field on $\mathbb{RP}^2$ is 1. Triangulate $\mathbb{RP}^2$ and compute its Euler characteristic.

Another nonorientable surface is the *Klein bottle*, which can be realized by identifying the opposite sides of a rectangle as shown:

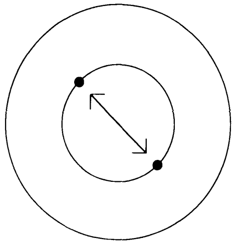


Exercise 8.13. What is the Euler characteristic of the Klein bottle?

The *Moebius band* is a nonorientable surface whose boundary is a circle:

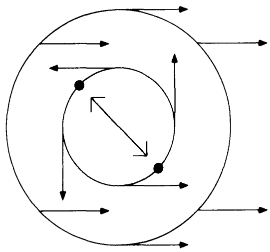


Or one may construct a Moebius band by taking an annulus, and identifying opposite points of one of the boundary circles:



Exercise 8.14. (a) Show that these two descriptions agree. (b) What do you get when you sew two Moebius bands together along their boundary circles?

Exercise 8.15. Find a vector field on the Moebius band with the boundary behavior shown and one singular point.



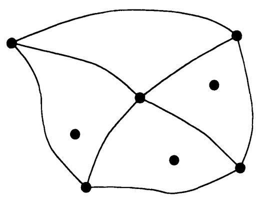
What is the index at the singular point?

Starting with any surface, one can cut out a disk, and paste back in a Moebius band, by identifying points on the boundary circles. This is called a *crosscap*.

Project 8.16. Investigate the surfaces that arise this way, especially the sums of indices of vector fields and the notion of Euler characteristic. What do you get when you do this to a sphere? What happens to the Euler characteristic of a surface when this is done to it? What is the Euler characteristic of the surface obtained by punching h disjoint disks from a sphere with g handles, and sewing h Moebius bands onto their boundaries. Can you realize the Klein bottle this way? Can two of these be homeomorphic, if you start with a different g and h? When?

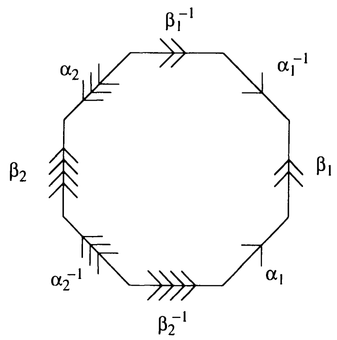
There is a beautiful proof of the Poincaré–Hopf theorem for a compact surface X that is oriented, using an inner product, varying con-

tinuously, on each tangent space. (If the surface is embedded in 3-space, one can identify each tangent space $T_P X$ as a subspace of $\mathbb{R}^3$, and use the standard inner product on $\mathbb{R}^3$.) To show that two vector fields V and W on X have the same sum of indices, triangulate X so all the singularities of V or W are inside triangles, and so that each triangle contains at most one singularity.



If P is a singular point in a triangle T, one can see that the number $\text{Index}_P V - \text{Index}_P W$ is the change in angle of the vector field $V - W$ around the boundary of T, divided by 2π; this change in angle is zero if there is no singularity in T. Adding over all the triangles, noting the cancellations along the edges, the theorem follows. (See Hopf (1983).)

There is a related proof, closer to our first proof for a torus. As we will see in Chapter 17, the surface X can be realized by identifying sides as indicated on a plane polygon R with $4g$ sides:



Exercise 8.17. Show that the resulting surface is a sphere with g handles.

Given vector fields V and W on X, one can find such a realization

of X so that none of the edges go through singularities of V or W. Then V and W determine vector fields $\widetilde{V}$ and $\widetilde{W}$ on the polygon, and

$$\sum_{P \in X} \operatorname{Index}_P V - \operatorname{Index}_P W = \sum_{P \in R} \operatorname{Index}_P \widetilde{V} - \operatorname{Index}_P \widetilde{W}$$
$$= \frac{1}{2\pi} (\text{change in angle of } \widetilde{V} - \widetilde{W} \text{ around } \partial R) = 0,$$

the last since the changes over identified edges of the boundary cancel.

We saw that the Euler characteristic is $2-2g$ by looking at the vector field of a fluid flowing down the surface, or the gradient of the height function for the surface sitting nicely in space. That picture also shows how one can build up the topology of X by looking at the portion of X whose height is at most h, and seeing how the topology changes as h increases. One sees that changes occur only when the height crosses the singularities, and that the change there is controlled by the indices of these singularities. This is the beginning of the beautiful subject of *Morse theory*. See Milnor (1963).

There is an important method for reducing some problems about nonorientable surfaces to the case of orientable surfaces. For a nonorientable surface X, there is an orientable surface $\widetilde{X}$ and a two-to-one mapping $p: \widetilde{X} \to X$, which is a local diffeomorphism. The two points in $\widetilde{X}$ over a point P in X correspond to the two ways to orient X near P. For example, if X is the projective plane, then $\widetilde{X}$ is the sphere. We will discuss this in more detail when we come to covering spaces, see §16a.

Problem 8.18. Show that a vector field V on X determines a vector field $\widetilde{V}$ on $\widetilde{X}$, and that the sum of the indices of $\widetilde{V}$ is twice the sum of the indices of V. Deduce that if X is any compact surface, the sum of indices of all vector fields with singularities on X is the same. If X is the surface constructed by sewing h Moebius bands to a sphere with g handles with h disks removed, show that $\widetilde{X}$ is a sphere with $2g+h-1$ handles.

PART V

COHOMOLOGY AND HOMOLOGY, II

In Chapter 9 the first homology group H_1U of a plane region U is computed for a plane region "with n holes." The notion of the integral of a closed $\mathscr{C}^\infty$ 1-form over any continuous path or any 1-chain is defined. We show that these integrals are the same over homologous 1-chains. This leads to useful methods for computing integrals and winding numbers, and relating the two notions; these are described in the third section.

The last section of Chapter 9, which is optional, takes a look at how these ideas are used in complex analysis. This is written as a (very) short course in complex analysis. It is self-contained, except for some calculations left as exercises, but in practice it will probably be most useful to those who have seen some of it before. For example, if you have seen Cauchy's formula and the residue theorem for regions such as disks and rectangles, this will show how the ideas of topology lead to the appropriate generalizations involving winding numbers. (This section includes all the analysis that will be needed when we study Riemann surfaces in Part X.)

The basic theme of the Mayer–Vietoris story in Chapter 10 is to find relations among the homology and cohomology groups of two open sets and their union and intersection. This possibility of comparing different homology groups H_k, for different k as well as different spaces, is a salient feature of algebraic topology. For cohomology, the beginnings of this story were seen in Chapter 5. This

chapter proves analogous results for homology, and then completes the cohomology story. This package of results, called the Mayer–Vietoris theorem, gives a powerful tool for calculating homology and cohomology groups.

It will be evident that the cohomology and homology groups behave in a similar, or more precisely, dual fashion. This duality will be made explicit in Part VIII. In fact, the proof of the full Mayer–Vietoris theorem for cohomology of plane domains will depend on this duality.

CHAPTER 9

Holes and Integrals

9a. Multiply Connected Regions

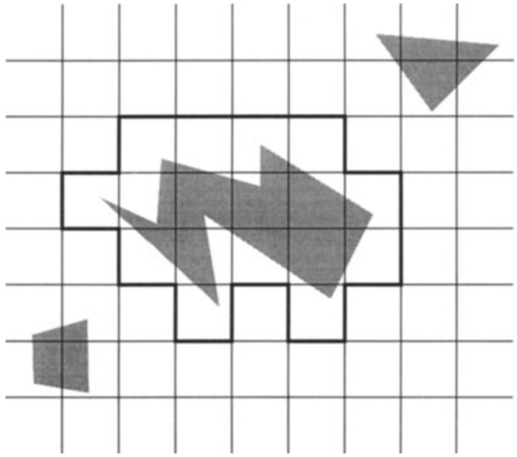
Let U be an open set in the plane, and let $A = \mathbb{R}^2 \setminus U$. We know that for a fixed closed chain γ in an open set U, the function $P \mapsto W(\gamma, P)$ is constant on connected components of $\mathbb{R}^2 \setminus \text{Supp}(\gamma)$. It is therefore constant on connected components of A, since each connected component of A is contained in some connected component of $\mathbb{R}^2 \setminus \text{Supp}(\gamma)$. If A is a connected component of $\mathbb{R}^2 \setminus U$, we write $W(\gamma, A)$ for the value of $W(\gamma, P)$ for all P in A, and call it the *winding number of* γ *around* A. Note that the connected components of a closed set such as $\mathbb{R}^2 \setminus U$ are closed, but they need not be path-connected (see Exercise 9.5).

Our goal is to calculate the first homology group of U, at least if this complement A is not too complicated. The idea is to find a closed path or 1-chain that "goes once" around each "piece" of A, and to show that these give a free basis for H_1U. The following lemma gives a precise statement and rigorous proof that this is possible:

Lemma 9.1. *Suppose A is a disjoint union of two closed sets B and C, with B bounded. Then there is a closed 1-chain γ in U such that $W(\gamma, P) = 1$ for all P in B and $W(\gamma, P) = 0$ for all P in C.*

Proof. By the compactness of B, there is a positive ε so that every point of B is at least distance ε away from any point of C. Take a

grid G so that its bounded (closed) rectangles cover B, but such that none of these rectangles meets both B and C, and so that none of the infinite rectangles meets B. This can be achieved by taking an infinite grid so that the distances between parallel lines is less than $\varepsilon/\sqrt{2}$, and letting G be the collection of lines that hit B.



The idea is to define γ to be the sum of the boundaries of the (closed) bounded rectangles in the grid that meet B:

$$\gamma = \sum_{R_k \cap B \neq \emptyset} \partial R_k .$$

Clearly γ is a closed 1-chain, since it is the sum of closed 1-chains. We claim next that γ is a 1-chain in U. To see this, suppose an edge σ occurs with a nonzero coefficient in γ, but that σ is not contained in U. We know that σ cannot meet C, since it is in a rectangle that meets B, and none of the rectangles meets both. So σ must meet B. But then each of the two (bounded) rectangles that σ separates meets B, so they occur in the displayed sum. Their contributions to σ therefore cancel, which shows that σ cannot occur in γ.

Next we show that $W(\gamma, P) = 1$ for P in B. Any point P in B is in one of the closed bounded rectangles, say R_l, and we let Q be a point in the interior of R_l. Using the fact that P and Q are in the same component of the complement of the support of γ, we have

$$W(\gamma, P) = W(\gamma, Q) = \sum_k W(\partial R_k, Q) = W(\partial R_l, Q) = 1.$$

Similarly, if $P \in C$, then $W(\gamma, P) = \Sigma_k W(\partial R_k, P) = 0$. □

A 1-chain γ having the property of the lemma is certainly not unique, but any other 1-chain with this property would have the same winding numbers around all points not in U, and it follows from Theorem 6.11 that it would be homologous to γ.

We next describe what it means for U to have "n holes." First we describe the "infinite" part of the complement A, denoted by A_∞, which will be a closed set with the property that winding numbers of closed 1-chains in U around points in A_∞ are always zero. In most examples it is obvious what A_∞ should be, but the general definition is a little complicated. Certainly A_∞ should contain any unbounded connected component of A; to assure that we get a closed set, we define A_∞ by the following:

$$A_\infty = \{P \in A: \text{for any } \varepsilon > 0 \text{ there is a connected subset } C \text{ of } A \text{ containing a point within distance } \varepsilon \text{ of } P \text{ and a point farther than } 1/\varepsilon \text{ from the origin}\}.$$

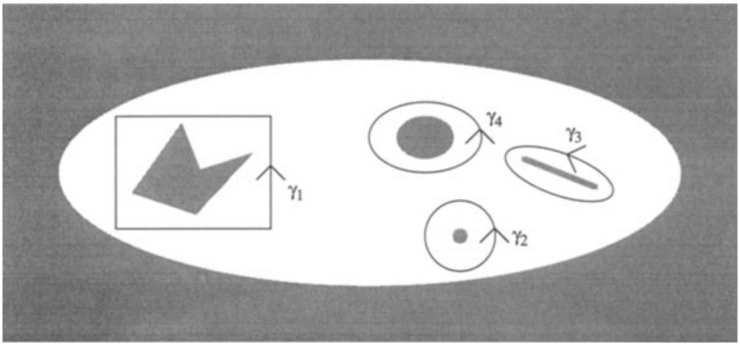
Lemma 9.2. *The set A_∞ is a closed subset of A, and $W(\gamma, P) = 0$ for all closed 1-chains γ in U and all points P in A_∞.*

Proof. If P is in the closure of A_∞, take a sequence P_n in A_∞ approaching P, and a connected set C_n as in the definition for P_n for some ε_n, with $\varepsilon_n \to 0$ as $n \to \infty$. It follows immediately that P is in A_∞. For any compact subset K of U, it follows from the definition that any point of A_∞ is in the unbounded component of $\mathbb{R}^2 \setminus K$. Note that the support of any 1-chain is compact. By Proposition 6.8 the winding number of any closed 1-chain around a point of A_∞ must be zero. □

Note that A_∞ can have more than one connected component, or it can be empty. Now suppose that

$$\mathbb{R}^2 \setminus U = A_1 \cup A_2 \cup \ldots \cup A_n \cup A_\infty,$$

where each A_i, $1 \le i \le n$, is a connected, closed, bounded, nonempty set, and these $n + 1$ sets are disjoint. Such an open set U is sometimes called *multiply connected*, or $(n + 1)$*-connected*. Not every U has such a complement, see Exercise 9.5.



By Lemma 9.1, for each i between 1 and n there is a closed chain γ_i in U so that $W(\gamma_i, A_i) = 1$, and $W(\gamma_i, A_j) = 0$ for $j \neq i$. Such 1-chains are not unique, but any other choices are homologous to these.

Proposition 9.3. *Any closed* 1*-chain* γ *on* U *is homologous to the* 1*-chain* $m_1\gamma_1 + \ldots + m_n\gamma_n$*, where* $m_i = W(\gamma, A_i)$*.*

Proof. With $m_i = W(\gamma, A_i)$, γ and $\Sigma m_i\gamma_i$ have the same winding numbers around all points not in U, and the result follows from Theorem 6.11. □

The integers $m_i = W(\gamma, A_i)$ are uniquely determined by the homology class γ, again by Theorem 6.11. In other words:

Corollary 9.4. *The homology classes of the closed chains* $\gamma_1, \ldots, \gamma_n$ *form a free basis of* H_1U*, giving an isomorphism* $H_1U \cong \mathbb{Z}^n$*.*

Exercise 9.5. (a) Let U be the set of points (x, y) such that $|y| < 1$ or $2n < x < 2n + 1$ for some integer n. Show that A_∞ is the union of an infinite number of connected components. (b) Let A be the union of the interval $\{(0, y)\colon 0 \leq y \leq 1\}$ and the set $\{(x, \sin(1/x))\colon x > 0\}$. Show that A is closed and connected, but not path-connected. (c) Let A be the union of the origin and the points $(1/n, 0)$ for n a positive integer. Show that each point of A is a connected component of A, but the origin has no neighborhood disjoint from the other components.

Exercise 9.6. Show that if $U \subset U'$ and U is n-connected, and U' is n'-connected, and $n > n'$, then there is no retract from U' onto U.

Problem 9.7. (a) Show that if U is the complement of the set $\mathbb{N}$ of nonnegative integers (identify n with $(n, 0)$), then H_1U is isomorphic

to the free abelian group on the points in $\mathbb{N}$. In particular, H_1U is not finitely generated. (b) Compute H_1U when U is the complement of the set in Exercise 9.5(c).

Problem 9.8. (a) If U is bounded, show that A_∞ is a connected component of $\mathbb{R}^2 \setminus U$. (b) If the plane is identified with the complement of the north pole in a sphere S (by stereographic projection, see §7c), show that the union of A_∞ and the north pole is the connected component of the complement of U in S that contains the north pole.

Problem 9.9. Generalize Corollary 9.4 as follows. Suppose U is any open set in the plane, and K is a compact subset of U that has n connected components $K_1, \ldots, K_n$. There is a homomorphism

$$H_1(U \setminus K) \to H_1U \oplus \mathbb{Z} \oplus \ldots \oplus \mathbb{Z} \cong H_1U \oplus \mathbb{Z}^n,$$

defined as follows. Given a closed 1-chain γ on $U \setminus K$, the first component of this map takes the class of γ in $H_1(U \setminus K)$ to the class of γ in H_1U, and the other components take the class of γ to the winding numbers of γ around the components $K_1, \ldots, K_n$. Show that this homomorphism is an isomorphism.

9b. Integration over Continuous Paths and Chains

We want to define the notion of the integral $\int_\gamma \omega$ of a closed $\mathscr{C}^\infty$ 1-form ω on an open set U over an arbitrary closed path γ in U. We cannot use calculus, but the idea we used for defining the winding number works perfectly well. If γ is defined on an interval $[a, b]$, subdivide the interval into $a = t_0 \le t_1 \le \ldots \le t_n = b$, so that each subinterval $[t_{i-1}, t_i]$ is mapped by γ into some open rectangle U_i contained in U. Such a subdivision exists by the Lebesgue lemma, since each point in the image of γ is contained in some such rectangle. The restriction of ω to U_i is exact (by Proposition 1.12), so we may find a $\mathscr{C}^\infty$ function f_i on U_i such that $df_i = \omega$ on U_i. Let $P_i = \gamma(t_i)$, $0 \le i \le n$. Define the integral $\int_\gamma \omega$ by

$$\int_\gamma \omega = (f_1(P_1) - f_1(P_0)) + (f_2(P_2) - f_2(P_1)) + \ldots + (f_n(P_n) - f_n(P_{n-1})).$$

The proof that this definition is independent of choices is almost the same as that of Proposition 3.1. Since the intersection of two

rectangles containing $\gamma([t_{i-1}, t_i])$ is connected, and f_i is unique up to adding a constant on a connected set, the sum is independent of choices of U_i and f_i. The rest of the proof is identical, by observing that refining a subdivision by adding a point doesn't change the answer.

If γ is a $\mathscr{C}^\infty$ path, it follows also from Proposition 1.16 that this definition of $\int_\gamma \omega$ agrees with that using calculus.

We extend the notion to all 1-chains $\gamma = n_1\gamma_1 + \ldots + n_r\gamma_r$, with γ_i paths, by linearity:

$$\int_\gamma \omega = n_1 \int_{\gamma_1} \omega + \ldots + n_r \int_{\gamma_r} \omega .$$

Note that the winding number is a special case of integral:

$$W(\gamma, P) = \int_\gamma \omega_P ,$$

where, if $P = (x_0, y_0)$,

$$\omega_P = \frac{1}{2\pi}\omega_{P,\vartheta} = \frac{1}{2\pi}\cdot\frac{-(y-y_0)\,dx + (x-x_0)\,dy}{(x-x_0)^2+(y-y_0)^2} .$$

Exercise 9.10. If $\omega = df$ is exact, and γ is a 1-chain with boundary $\partial\gamma = \Sigma m_j P_j$, show that $\int_\gamma \omega = \Sigma m_j f(P_j)$. In particular, $\int_\gamma df = 0$ if γ is closed.

Proposition 9.11. *If γ and δ are homologous 1-chains in U, then*

$$\int_\gamma \omega = \int_\delta \omega .$$

Proof. The proof is the same as the proof of Lemma 6.7, which refers to Theorem 3.6. One simply changes winding numbers to integrals, sectors $U_{i,j}$ to open rectangles, and angle functions ϑ_{ij} to arbitrary functions $f_{i,j}$ with $df_{i,j} = \omega$ on $U_{i,j}$. □

It follows in particular that the integral of ω is the same over two paths that are homotopic closed paths, or homotopic paths with the same endpoints. It also follows that the integrals are unchanged by the reparametrization of paths.

Corollary 9.12. *Given two closed 1-chains γ and δ on U, the following are equivalent:*

(1) *$W(\gamma, P) = W(\delta, P)$ for all P not in U;*
(2) *$\int_\gamma \omega = \int_\delta \omega$ for all closed 1-forms ω on U; and*

(3) $\gamma - \delta$ *is a boundary* 1-*chain in* U.

Proof. The implications (3)$\Rightarrow$(2)$\Rightarrow$(1) have just been seen, and (1)$\Leftrightarrow$(3) is Theorem 6.11. □

Corollary 9.13. *Let* U *be any open set in the plane. Then the following are equivalent*:

(1) $W(\gamma, P) = 0$ *for all closed chains* γ *in* U *and all* $P \notin U$;
(2) $\int_\gamma \omega = 0$ *for all closed chains* γ *in* U *and all closed* 1-*forms* ω *on* U;
(3) *every closed* 1-*chain is a boundary*: $H_1 U = 0$; *and*
(4) *every closed* 1-*form* ω *on* U *is exact*: $H^1 U = 0$.

These conditions hold whenever U *is* 1-*connected, i.e.*, $\mathbb{R}^2 \setminus U = A_\infty$.

Proof. The equivalence of the first three conditions follows from the preceding corollary. We have (4)$\Rightarrow$(2) by Exercise 9.10. Conversely, (2) implies that $\int_\gamma \omega = \int_\delta \omega$ whenever γ and δ are segmented paths with the same endpoints, and ω is any closed 1-form; we saw in Chapter 1 that this makes ω exact, which is (4). □

We know that homotopic closed paths are homologous.

Problem 9.14. (a) Show that homologous closed paths must be homotopic when U is an open rectangle or any convex open set. (b) Do the same when U is the complement of a point, or an annulus

$$U = \{(x, y): {r_1}^2 < (x - x_0)^2 + (y - y_0)^2 < {r_2}^2\}.$$

(c) *Challenge*. What if U is the complement of two points?

Problem 9.15. Suppose $\gamma = \Sigma n_i \gamma_i$ is a closed 1-chain in an open set U such that each γ_i is a $\mathscr{C}^\infty$ path. If γ is homologous to zero, show that one can write $\gamma = \Sigma m_j(\partial \Gamma_j)$, where each Γ_j is a $\mathscr{C}^\infty$ map from the unit square to U.

Problem 9.16. If X is any closed, connected subset of the plane, and U is a bounded connected component of $\mathbb{R}^2 \setminus X$, show that every closed 1-form on U is exact.

Problem 9.17. (a) If $\omega_t = p(x, y, t)\,dx + q(x, y, t)\,dy$ is a continuously varying family of closed $\mathscr{C}^\infty$ 1-forms on U, i.e., the functions p and q are continuous on $U \times [a, b]$, show that the function $t \rightarrow \int_\gamma \omega_t$ is a continuous function of t. (b) Use this to give another proof of Proposition 3.16.

9c. Periods of Integrals

With U an $(n+1)$-connected region as above, and ω any closed 1-form on U, define the *period* (or *module of periodicity*) of ω *around* A_i to be the integral of ω along γ_i, with γ_i as defined just before Proposition 9.3. By Proposition 9.11, this is independent of choice of γ_i. Denote this period by $\mathfrak{p}_i(\omega)$ or $\mathfrak{p}(\omega, A_i)$:

$$\mathfrak{p}_i(\omega) = \mathfrak{p}(\omega, A_i) = \int_{\gamma_i} \omega .$$

These numbers determine integrals of ω along any closed path, provided one knows the winding number of the path around each A_i:

Proposition 9.18. *For any closed* 1*-chain* γ *and closed* 1*-form* ω,

$$\int_\gamma \omega = W(\gamma, A_1)\,\mathfrak{p}_1(\omega) + W(\gamma, A_2)\,\mathfrak{p}_2(\omega) + \ldots + W(\gamma, A_n)\,\mathfrak{p}_n(\omega) .$$

Proof. Since γ and $\Sigma_i W(\gamma, A_i) \cdot \gamma_i$ are homologous, this follows from Proposition 9.11. □

Applying Corollary 9.12, we have:

Corollary 9.19. *A closed* 1*-form* ω *on* U *is exact if and only if all of its periods* $\mathfrak{p}_i(\omega)$ *are zero.*

For integrals along paths that are not closed, if γ and δ are two paths with the same endpoints, applying the proposition to $\gamma - \delta$, we have

$$\int_\gamma \omega - \int_\delta \omega = m_1\,\mathfrak{p}_1(\omega) + m_2\,\mathfrak{p}_2(\omega) + \ldots + m_n\,\mathfrak{p}_n(\omega) ,$$

with $m_1, \ldots, m_n$ integers. This means that the integral is determined up to adding integral combinations of the periods. For example, when $U = \mathbb{R}^2 \setminus \{P\}$ and $\omega = \omega_{P,\vartheta}$ (see Problem 2.10), there is only one period, which is 2π, and we recover the fact that the integral is determined up to integral multiples of 2π.

Exercise 9.20. Compute the integral $\int_\gamma \omega$, where ω is the 1-form

$$\omega = \sum_{n=1}^{17} \frac{-y\,dx + (x-n)\,dy}{(x-n)^2 + y^2} ,$$

and $\gamma(t) = (t\cos(t), t\sin(t))$, $0 \le t \le 6\pi$.

Exercise 9.21. Show that, given U as above, for any real numbers $\mathfrak{p}_1, \ldots, \mathfrak{p}_n$ there is a closed 1-form ω with these periods.

This means that the linear mapping from the vector space of closed 1-forms on U to $\mathbb{R}^n$, $\omega \mapsto (\mathfrak{p}_1(\omega), \ldots, \mathfrak{p}_n(\omega))$, is surjective, and by Corollary 9.19 the kernel is the space of exact 1-forms. This sets up an isomorphism (see §C1)

$$\{\text{closed 1-forms on } U\}/\{\text{exact 1-forms on } U\} \cong \mathbb{R}^n,$$

i.e., the De Rham group H^1U is an n-dimensional vector space. If P_i is any point in A_i, $1 \leq i \leq n$, the classes $[\omega_{P_i}]$ form a basis for $H^1(U)$.

Problem 9.22. (a) Suppose $n = 2$, and the periods of ω are $\mathfrak{p}_1$ and $\mathfrak{p}_2$. Let P and Q be two fixed points in U. Show that if the periods are not zero, and the ratio $\mathfrak{p}_1/\mathfrak{p}_2$ is rational, there is a number r so that if γ and δ are any two paths from P to Q, then $\int_\gamma \omega - \int_\delta \omega$ is an integer times r. (b) Show that, if $\mathfrak{p}_1/\mathfrak{p}_2$ is not rational, and U is connected, there is no such r.

9d. Complex Integration

The plane $\mathbb{R}^2$ can be identified with the complex numbers $\mathbb{C}$, the pair (x, y) being identified with $z = x + iy$. Functions on open sets in the plane will be written as functions of z.

A *complex 1-form* ω on an open set U in the plane is given by a pair of ordinary (real) 1-forms ω_1 and ω_2, written in the form

$$\omega = \omega_1 + i\,\omega_2.$$

Define $d\omega = d\omega_1 + i\,d\omega_2$, and for a function $f = u + iv$, with u and v real-valued functions on U, set $df = du + i\,dv$. The form ω is *closed* if $d\omega = 0$, and *exact* if $\omega = df$. For example, we have the 1-form dz defined by $dz = dx + i\,dy$. A complex 1-form can by multiplied by a complex-valued function: if ω is as above, and $f = u + iv$, then $f \cdot \omega$ is the complex 1-form

$$f \cdot \omega = (u + iv) \cdot (\omega_1 + i\omega_2) = (u\omega_1 - v\omega_2) + i(u\omega_2 + v\omega_1).$$

Exercise 9.23. (a) If $f = u + iv$, with u and v $\mathscr{C}^\infty$ functions, show that the 1-form $f(z)\,dz$ is closed if and only if u and v satisfy the Cauchy–Riemann equations:

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y} \quad \text{and} \quad \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x}.$$

(b) Show that if the conditions in (a) are satisfied, then the complex derivative $f'(a) = \lim_{z \to a}(f(z) - f(a))/(z - a)$ exists at each point a in U.

Let us call f *analytic* if its real and imaginary parts satisfy the Cauchy–Riemann equations. For example, if f is locally expandable in a power series, then f is analytic, so $f(z)\,dz$ is closed. Indeed, if $f(z) = \sum_{n=0}^{\infty} a_n(z - z_0)^n$ in a disk around z_0, then $f = dg$ on that disk, with

$$g(z) = \sum_{n=0}^{\infty} \frac{a_n}{n+1}(z - z_0)^{n+1}.$$

For an example that is closed but not exact, consider dz/z on the complement of the origin:

$$\begin{aligned} \frac{dz}{z} &= \frac{dx + i\,dy}{x + iy} = \frac{(x - iy)(dx + i\,dy)}{x^2 + y^2} \\ &= \frac{x\,dx + y\,dy}{x^2 + y^2} + i\frac{-y\,dx + x\,dy}{x^2 + y^2} = d(\log(r)) + i\omega_\vartheta \end{aligned}$$

Similarly, $dz/(z - a) = d(\log(|z - a|)) + i\omega_{a,\vartheta}$, with $\omega_{a,\vartheta}$ as defined in Chapter 2.

If γ is a path or chain in U, the integral of $\omega = \omega_1 + i\omega_2$ along γ is defined by

$$\int_\gamma \omega = \int_\gamma \omega_1 + i\int_\gamma \omega_2.$$

For example, if $\gamma(t) = a + r \cdot e^{it}$, $0 \le t \le 2\pi$, is a circle around the point a, then

$$\int_\gamma \frac{dz}{z - a} = i\int_\gamma \omega_{a,\vartheta} = 2\pi i.$$

In general, if γ is any chain not containing a in its support, we see similarly that

$$\int_\gamma \frac{dz}{z - a} = i\int_\gamma \omega_{a,\vartheta} = 2\pi i \cdot W(\gamma, a).$$

The main fact about complex integrals is

Theorem 9.24 (Cauchy Integral Theorem). *If γ is a closed chain in U whose winding number around any point not in U is zero, and f is an analytic function in U, then for any a in U that is not in the*

support of γ,

$$W(\gamma, a)\cdot f(a) \;=\; \frac{1}{2\pi i}\int_\gamma \frac{f(z)}{z-a}\,dz\,.$$

Proof. Look at $F(z) = (f(z) - f(a))/(z-a)$, which is analytic on $U \setminus \{a\}$. Using the formula proved just before the statement of the theorem, the displayed formula is equivalent to the formula $\int_\gamma F(z)\,dz = 0$. Let $\delta_r(t) = a + re^{2\pi i t}$, $0 \le t \le 1$, with r small enough so the disk of radius r around a is contained in U. Let $n = W(\gamma, a)$. Since $W(\gamma, P) = n\cdot W(\delta_r, P)$ for all $P \notin U \setminus \{a\}$, we know by Corollary 9.12 that $\int_\gamma F(z)\,dz = n\cdot\int_{\delta_r} F(z)\,dz$, so it suffices to prove that $\lim_{r\to 0}\int_{\delta_r} F(z)\,dz = 0$. But by Exercise 9.23(b), $F(z)$ has a limit as z approaches a, and the fact that $\int_{\delta_r} F(z)\,dz$ approaches zero as the radius goes to zero follows easily (see Exercise B.8). □

The simplest form of Cauchy's formula is when γ is a circle about a. Since the right side of Cauchy's formula is expandable in a power series in a neighborhood of a, this implies in particular the fact that any analytic function is locally expandable in a power series.

Exercise 9.25. (a) Let U be an open set containing two concentric circles and the region between them. For a in the region between the circles, show that

$$f(a) \;=\; \frac{1}{2\pi i}\int_{\gamma_r} \frac{f(z)}{z-a}\,dz - \frac{1}{2\pi i}\int_{\gamma_\varepsilon} \frac{f(z)}{z-a}\,dz\,,$$

where γ_r and γ_ε are counterclockwise paths around the larger and smaller circle. (b) Deduce Riemann's theorem on removable singularities: if f is analytic and bounded in a punctured neighborhood of a point b, then f extends to an analytic function in a full neighborhood of b.

A related application is to the general residue theorem. If f is analytic in a punctured neighborhood of a point a (i.e., in some $U \setminus \{a\}$, for U a neighborhood of a), the *residue* of f at a, denoted $\mathrm{Res}_a(f)$, is defined by

$$\mathrm{Res}_a(f) \;=\; \frac{1}{2\pi i}\int_\delta f(z)\,dz\,,$$

where δ is a small counterclockwise circle around a. By the Cauchy integral theorem, this is independent of the circle chosen.

Exercise 9.26. If f is given by a converging series $\Sigma_{n=-m}^{\infty} c_n(z-a)^n$ in

a punctured neighborhood of a (i.e., f has at most a *pole* at a), show that $\mathrm{Res}_a(f) = c_{-1}$.

Theorem 9.27 (Residue Theorem). *If f is analytic in $U \setminus \{a_1, \ldots, a_r\}$, and γ is a closed 1-chain in $U \setminus \{a_1, \ldots, a_r\}$ such that $W(\gamma, P) = 0$ for all P not in U, then*

$$\frac{1}{2\pi i}\int_\gamma f(z)\,dz = \sum_{i=1}^{r} W(\gamma, a_i)\cdot \mathrm{Res}_{a_i}(f).$$

Proof. Take disjoint small circles δ_i around a_i. Then, with $n_i = W(\gamma, a_i)$, γ and $\Sigma n_i\delta_i$ have the same winding number around all points not in $U \setminus \{a_1, \ldots, a_r\}$. It follows from Corollary 9.12 that the integral of the closed 1-form $f(z)\,dz$ around γ and around $\Sigma n_i\delta_i$ gives the same answer. □

Problem 9.28. Extend the residue theorem to allow f to be analytic outside any discrete set S in U (i.e., each point in U has a neighborhood containing at most one point of S).

Exercise 9.29. Suppose U is an $(n+1)$-connected open set as in the first section, with $\mathbb{R}^2 \setminus U = A_1 \cup \ldots \cup A_n \cup A_\infty$, and f is analytic on U; let $\mathfrak{p}_i$ be the period of $f(z)\,dz$ around A_i. Show that, for a closed chain γ on U,

$$\int_\gamma f(z)\,dz = W(\gamma, A_1)\,\mathfrak{p}_1 + \ldots + W(\gamma, A_n)\,\mathfrak{p}_n.$$

Exercise 9.30. Deduce the Cauchy Integral Theorem directly from the Residue Theorem.

Suppose now f is *meromorphic* in U, i.e., near every point a in U one can write $f(z) = (z-a)^m \cdot h(z)$, with h analytic at a, $h(a) \neq 0$, and m an integer. The *order of f at a*, denoted $\mathrm{ord}_a(f)$, is defined to be this integer m.

Exercise 9.31. Show that, with f and h as above, if δ is the boundary of a disk about a such that h is nowhere zero in the disk, then

$$\mathrm{ord}_a(f) = W(f\circ\delta, 0).$$

Theorem 9.32 (Argument Principle). *Suppose f is meromorphic in*

U, and γ is a closed 1-chain in U not passing through any zero or pole of f, such that $W(\gamma, P) = 0$ for all P not in U. Then

$$W(f \circ \gamma, 0) = \sum_a W(\gamma, a) \cdot \mathrm{ord}_a(f),$$

where the sum is over the (finitely many) zeros or poles a for which $W(\gamma, a) \neq 0$.

Proof. If $\{a_1, \ldots, a_r\}$ are the zeros and poles around which γ has a nonzero winding number (see Problem 9.28), take a small circle δ_i around a_i as in the proof of the Residue Theorem. Applying Proposition 6.15 to the mapping $f\colon U \setminus \{a_1, \ldots, a_r\} \to U' = \mathbb{C} \setminus \{0\}$, we get the formula $W(f \circ \gamma, 0) = \sum_{i=1}^{r} W(\gamma, a_i) \cdot W(f \circ \delta_i, 0)$. □

The following problems give the analytic interpretation, and a typical application, of the Argument Principle:

Problem 9.33. (a) If f is meromorphic at a, show that

$$\mathrm{ord}_a(f) = \mathrm{Res}_a\left(\frac{f'}{f}\right).$$

(b) Under the conditions of the Argument Principle, use the Residue Theorem to show that

$$\frac{1}{2\pi i}\int_\gamma \frac{f'(z)}{f(z)}\,dz = \sum_a W(\gamma, a) \cdot \mathrm{ord}_a(f).$$

(c) Show directly that

$$\frac{1}{2\pi i}\int_\gamma \frac{f'(z)}{f(z)}\,dz = W(f \circ \gamma, 0)$$

by interpreting the integrand as $d(\log(f(z)) = d(\log|f(z)|) + i \cdot d(\arg(f(z)))$, with $\log|f(z)|$ a well-defined function, and $\arg(f(z))$ the multivalued angle function.

Problem 9.34. (a) Let U, f, and γ be as in the Argument Principle. If g is analytic function on U, show that

$$\frac{1}{2\pi i}\int_\gamma g(z)\frac{f'(z)}{f(z)}\,dz = \sum_a W(\gamma, a) \cdot g(a) \cdot \mathrm{ord}_a(f).$$

(b) Suppose the restriction of f to a closed disk D around z_0 in U is one-to-one, and let γ be the boundary of D. Show that the function

that takes w to $(1/2\pi i)\int_\gamma z f'(z)/(f(z) - w)\,dz$ defines an inverse function to f in a neighborhood of $f(z_0)$.

The following is Rouché's theorem:

Problem 9.35. Suppose the chain γ is homologous to zero in U, and suppose f and g are analytic functions in U such that

$$|f(z) - g(z)| < |f(z)| + |g(z)|$$

for all z in the support of γ. Show that

$$\sum_a W(\gamma, a) \cdot \operatorname{ord}_a(f) = \sum_a W(\gamma, a) \cdot \operatorname{ord}_a(g).$$

In particular, if the winding number of γ is 1 for points a in some region V, and the winding number is zero elsewhere, then f and g have the same number of zeros in V, counting multiplicities.

CHAPTER 10

Mayer–Vietoris

10a. The Boundary Map

For open sets U and V in the plane (or any topological space) we will define a homomorphism

$$\partial\colon H_1(U \cup V) \to H_0(U \cap V),$$

called the *boundary map*.[5] To do this we need a lemma.

Lemma 10.1. *If γ is a 1-cycle on $U \cup V$, there are 1-chains γ_1 on U and γ_2 on V such that $\gamma_1 + \gamma_2$ is homologous to γ on $U \cup V$.*

Proof. We know from Lemma 6.4(b) that if a path is subdivided, it is homologous to the sum of the paths into which it is divided. By the Lebesgue lemma, each path occurring in γ can be subdivided so that the image of each piece is in U or in V. So γ is homologous to a sum $\sum n_i\tau_i$, where each τ_i is a path in U or in V (or both). Then γ_1 can be taken to be the sum of those $n_i\tau_i$ for which τ_i is a path in U, and γ_2 can be the sum of the others. □

Construction of the boundary map $\partial\colon H_1(U \cup V) \to H_0(U \cap V)$. Recall that $H_1(U \cup V) = Z_1(U \cup V)/B_1(U \cup V)$ is the group of 1-cycles,

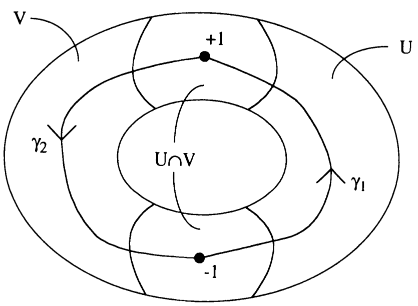
[5] This symbol ∂ is certainly overworked in this subject. So far we have used it for the boundaries of 1-chains and for maps of rectangles. Rather than introducing different notations for the different uses, we try for clarity by saying in each case to what sort of object the "∂" is being applied.

modulo the subgroup of 1-boundaries, on $U \cup V$. And $H_0(U \cap V) = Z_0(U \cap V)/B_0(U \cup V)$ is the group of 0-cycles modulo 0-boundaries on $U \cap V$. We will write $[\gamma]$ for the homology class in $H_1(U \cup V)$ defined by a 1-cycle γ on $U \cup V$, and we write $[\zeta]$ in $H_0(U \cap V)$ for the class defined by a zero cycle ζ on $U \cap V$.

Given a homology class α in $H_1(U \cup V)$, by the lemma, we may choose a 1-cycle that represents α and has the form $\gamma_1 + \gamma_2$, where γ_1 is a 1-chain on U and γ_2 is a 1-chain on V; in symbols, $[\gamma_1 + \gamma_2] = \alpha$. Since the boundary of this 1-cycle is zero, we have $\partial(\gamma_1) = -\partial(\gamma_2)$. This 0-cycle $\partial(\gamma_1) = -\partial(\gamma_2)$ is a 0-cycle on U and on V, so it is a 0-cycle on $U \cap V$. We define $\partial\colon H_1(U \cup V) \to H_0(U \cap V)$ by sending α to the class of this 0-cycle:

$$\partial(\alpha) = \partial([\gamma_1 + \gamma_2]) = [\partial(\gamma_1)] = -[\partial(\gamma_2)].$$

Note that, although the zero cycle $\partial(\gamma_1) = -\partial(\gamma_2)$ is a boundary on U and on V, it need not be a boundary on $U \cap V$:

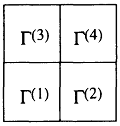


The hardest part of our task is to show that this class $[\partial\gamma_1]$ is well defined.

Lemma 10.2. *The class of $\partial\gamma_1$ in $H_0(U \cap V)$ is independent of the choice of γ_1 and γ_2.*

Proof. Before giving the proof, we need a more canonical way to subdivide our 1-chains, by cutting each one exactly in half. For any path $\gamma\colon [0, 1] \to U$, define the 1-chain $S(\gamma)$ by the formula $S(\gamma) = \sigma + \tau$, where σ and τ are the restrictions of γ to the two halves of the interval, but rescaled as in Lemma 6.4(b). Extend this operator S linearly to all 1-chains $\gamma = \Sigma n_i\gamma_i$ by setting $S(\gamma) = \Sigma n_i S(\gamma_i)$. It follows from this definition that the boundary of the 1-chain $S(\gamma)$ is the same as the boundary of γ.

If Γ: $[0, 1] \times [0, 1] \to U$ is a map of the unit square, we can similarly subdivide Γ into four pieces $\Gamma^{(1)}$, $\Gamma^{(2)}$, $\Gamma^{(3)}$, and $\Gamma^{(4)}$ as indicated:



Each restriction is rescaled to be a map from $[0, 1] \times [0, 1]$ to U. In formulas,

$$\Gamma^{(1)}(t, s) = \Gamma(1/2t, 1/2s), \qquad \Gamma^{(2)}(t, s) = \Gamma(1/2 + 1/2t, 1/2s),$$

$$\Gamma^{(3)}(t, s) = \Gamma(1/2t, 1/2 + 1/2s), \qquad \Gamma^{(4)}(t, s) = \Gamma(1/2 + 1/2t, 1/2 + 1/2s).$$

It is clear from the picture, and easy to verify from the formulas, that this subdivision is compatible with taking the boundary, i.e.,

$$S(\partial\Gamma) = \partial\Gamma^{(1)} + \partial\Gamma^{(2)} + \partial\Gamma^{(3)} + \partial\Gamma^{(4)},$$

the inside boundaries canceling as usual.

Now we prove the lemma. Suppose the class α is also represented by the cycle $\gamma_1' + \gamma_2'$ for 1-chains γ_1' on U and γ_2' on V. Since both sums represent γ, we know that $(\gamma_1 + \gamma_2) - (\gamma_1' + \gamma_2')$ is a boundary on $U \cup V$, so we can write

$$(\gamma_1 + \gamma_2) - (\gamma_1' + \gamma_2') = \Sigma n_i \partial\Gamma_i$$

for some maps Γ_i from rectangles to $U \cup V$. We must show that $[\partial\gamma_1] = [\partial\gamma_1']$ in $H_0(U \cap V)$, i.e., that $\partial\gamma_1 - \partial\gamma_1'$ is a boundary of some 1-chain on $U \cap V$. We apply the subdivision operator S to each side of the displayed equation. On the left, each of the four 1-chains is replaced by another with the same support and the same boundary. On the right, each $\partial\Gamma_i$ is replaced by a sum of the boundaries of the four subdivisions of Γ_i. So we have an equation of the same form, but with all the Γ_i's cut into quarters. The operator S can be applied again, which subdivides each of these quarters into quarters, and so on, dividing the smaller squares into quarters. When applied p times, we have an equation

$$S^p(\gamma_1) + S^p(\gamma_2) - S^p(\gamma_1') - S^p(\gamma_2') = \Sigma n_i S^p(\partial\Gamma_i).$$

By the Lebesgue lemma, the restrictions of Γ_i to small enough portions of the rectangles must be mapped into U or into V. It follows that for some large p the right side of this equation can be written in the form $\tau_1 + \tau_2$, where τ_1 is a 1-boundary on U and τ_2 is a 1-boundary

on V. This gives an equation of 1-chains:

$$S^p(\gamma_1) - S^p(\gamma_1') - \tau_1 \;=\; S^p(\gamma_2') - S^p(\gamma_2) + \tau_2 .$$

The left side is a 1-chain on U, the right side a 1-chain on V, so the 1-chain they define is a 1-chain on $U \cap V$. And the boundary of this chain is

$$\partial(S^p(\gamma_1) - S^p(\gamma_1') - \tau_1) \;=\; \partial\gamma_1 - \partial\gamma_1' - \partial\tau_1 \;=\; \partial\gamma_1 - \partial\gamma_1' ,$$

which shows that $\partial\gamma_1 - \partial\gamma_1'$ is a 0-boundary on $U \cap V$, and completes the proof. □

Lemma 10.3. *The boundary operator* $\partial \colon H_1(U \cup V) \to H_0(U \cap V)$ *is a homomorphism of abelian groups.*

Proof. This follows readily from the definition. For if α is represented by $\gamma_1 + \gamma_2$, and α' is represented by $\gamma_1' + \gamma_2'$, with γ_1 and γ_1' 1-chains on U and γ_2 and γ_2' 1-chains on V, then $\alpha \pm \alpha'$ is represented by $(\gamma_1 \pm \gamma_1') + (\gamma_2 \pm \gamma_2')$, so

$$\partial(\alpha \pm \alpha') \;=\; [\partial(\gamma_1 \pm \gamma_1')] \;=\; [\partial(\gamma_1)] \pm [\partial(\gamma_1')] \;=\; \partial\alpha \pm \partial\alpha' . \quad \square$$

10b. Mayer–Vietoris for Homology

If $U_1 \subset U_2$, we saw in Chapter 6 that there are homomorphisms from H_0U_1 to H_0U_2 and from H_1U_1 to H_1U_2. Given two open sets U and V we therefore have a diagram

$$\begin{array}{ccccccccc}
 & H_1U & & & & H_0U & \\
\nearrow & & \searrow & & \nearrow & & \searrow \\
H_1(U\cap V) & & H_1(U\cup V) & \xrightarrow{\;\partial\;} & H_0(U\cap V) & & H_0(U\cup V) . \\
\searrow & & \nearrow & & \searrow & & \nearrow \\
 & H_1V & & & & H_0V &
\end{array}$$

The Mayer–Vietoris story gives the relations among all these groups and homomorphisms. This can be described in a series of assertions, moving from right to left in the diagram.

MV(i). *Any element in* $H_0(U \cup V)$ *is the sum of the images of an element in* H_0U *and an element in* H_0V.

Proof. Any 0-chain that represents an element in $H_0(U \cup V)$ can be written as a sum of a 0-chain in U and a 0-chain in V. □

MV(ii). *An element in H_0U and an element in H_0V have the same image in $H_0(U \cup V)$ if and only if they come from some element in $H_0(U \cap V)$.*

Proof. Let b and c be 0-chains representing elements in H_0U and H_0V. If they have the same image in $H_0(U \cup V)$, there is a 1-chain γ on $U \cup V$ with $\partial\gamma = b - c$. By subdividing the paths in γ to be sufficiently small, so that the image of each subdivided path is contained in U or in V, we may find 1-chains γ_1 on U and γ_2 on V so that $\partial(\gamma_1 + \gamma_2) = \partial\gamma$. Then $b - \partial\gamma_1 = c + \partial\gamma_2$, and the left side of this is a 0-chain on U, and the right side is a 0-chain on V, so $a = b - \partial\gamma_1 = c + \partial\gamma_2$ is a 0-chain on $U \cap V$. Since a is homologous to b on U and to c on V, the class in $H_0(U \cap V)$ represented by a maps to the classes represented by b in H_0U and c in H_0V. The converse is immediate from the definitions. □

MV(iii). *An element in $H_0(U \cap V)$ maps to zero in H_0U and in H_0V if and only if it is the image by ∂ of some element in $H_1(U \cup V)$.*

Proof. We saw during the definition of the boundary ∂ that a boundary always has this form. Conversely, if $\zeta = \partial\gamma_1$ and $\zeta = \partial\gamma_2$ for 1-chains γ_1 on U and γ_2 on V, then $[\zeta] = \partial(\alpha)$, where α is the class represented by the 1-cycle $\gamma_1 - \gamma_2$. □

MV(iv). *An element in $H_1(U \cup V)$ maps to zero in $H_0(U \cap V)$ if and only if it is the sum of an element coming from H_1U and an element coming from H_1V.*

Proof. If $\alpha = [\gamma_1 + \gamma_2]$ for some 1-cycle γ_1 on U and some 1-cycle γ_2 on V, we may take γ_1 and γ_2 for the 1-chains in the definition of the boundary, so $\partial(\alpha) = [\partial\gamma_1] = 0$. Conversely, if α is in the kernel of ∂, and we write $\alpha = [\gamma_1 + \gamma_2]$, with γ_i as in the definition of the boundary, then $\partial\gamma_1 = -\partial\gamma_2$ is a 0-boundary on $U \cap V$, so $\partial\gamma_1 = \partial\tau$ for some 1-chain τ on $U \cap V$. Therefore,

$$\alpha = [(\gamma_1 - \tau) + (\gamma_2 + \tau)],$$

and $\gamma_1 - \tau$ is a 1-cycle on U and $\gamma_2 + \tau$ is a 1-cycle on V. □

MV(v). *An element in H_1U and an element in H_1V have the same image in $H_1(U \cup V)$ if and only if they come from some element in $H_1(U \cap V)$.*

Proof. Let β and γ be 1-cycles on U and V representing the two elements. If they have the same image in $H_1(U \cup V)$, there is an equation

$$\beta - \gamma = \Sigma n_i \partial \Gamma_i$$

for some maps Γ_i from $[0, 1] \times [0, 1]$ to $U \cup V$. We apply the subdividing operator S, which was introduced in the proof of Lemma 10.2, to this equation. As in that lemma, if S is applied sufficiently many times, the images of the subdivided rectangles will be contained in U or in V, and we will have an equation

$$S^p\beta - S^p\gamma = \Sigma n_i S^p(\partial \Gamma_i) = \delta_1 + \delta_2,$$

where δ_1 is a 1-boundary on U and δ_2 is a 1-boundary on V. Lemma 6.4(b) proves that S takes a 1-cycle to a 1-cycle that is homologous to it. So $S^n\beta - \delta_1 = S^p\gamma + \delta_2$ is a 1-cycle on $U \cap V$ that is homologous to β on U and to γ on V, as required. Again, the converse is obvious. □

In the case of open sets in the plane, there is one more assertion that completes the Mayer–Vietoris story.

MV(vi). *If U and V are open subsets of the plane, an element in $H_1(U \cap V)$ is zero if and only if its images in $H_1 U$ and in $H_1 V$ are zero.*

Proof. If γ is a 1-cycle representing the element in $H_1(U \cap V)$, if its images are zero in $H_1 U$ and $H_1 V$, then the winding number $W(\gamma, P)$ is zero for all P not in U and all P not in V. This means that $W(\gamma, P)$ vanishes for all P not in $U \cap V$, and by Theorem 6.11 this implies that γ is homologous to zero on $U \cap V$. □

Exercise 10.4. If $H_1 U = 0$ and $H_1 V = 0$, show that the kernel of ∂ is zero. If in addition $U \cap V$ is connected, show that $H_1(U \cup V) = 0$.

Exercise 10.5. If U and V are connected, show that the image of ∂ consists of all classes in $H_0(U \cap V)$ of degree zero.

Exercise 10.6. Show that if U and V are connected, and $H_1(U \cup V) = 0$, then $U \cap V$ is also connected.

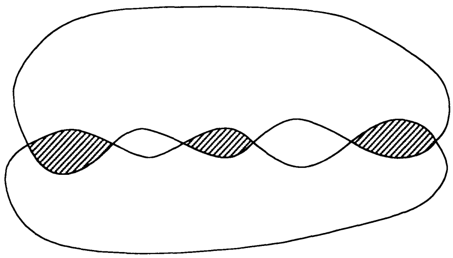
Exercise 10.7. If $U \cap V$ is connected, and $H_1(U \cap V) = 0$, show that $H_1(U \cup V)$ is isomorphic to the direct sum of $H_1 U$ and $H_1 V$.

Exercise 10.8. If the homomorphism from $H_1(U \cap V)$ to $H_1 V$ is in-

jective, show that the homomorphism from H_1U to $H_1(U \cup V)$ is also injective.

Exercise 10.9. Suppose A and B are disjoint closed subsets in a space X, and $H_1X = 0$. Show that if $X \setminus A$ and $X \setminus B$ are path-connected, then $X \setminus (A \cup B)$ is also path-connected.

Let us use Mayer–Vietoris to work out an example. We start with a familiar situation, with U and V as simple as possible, which arises when we take the union of two open sets each diffeomorphic to disks.



Suppose U and V are connected, with $H_1U = 0 = H_1V$. Suppose $U \cap V$ has $m \geq 1$ connected components. We claim that $H_1(U \cup V)$ is a free abelian group with $m - 1$ generators. By MV(iv), the boundary map ∂ is injective, and by MV(iii), the image of ∂ consists of elements of $H_0(U \cap V)$ that restrict to zero in $H_0U = \mathbb{Z}$ and $H_0V = \mathbb{Z}$. Now $H_0(U \cap V)$ is the free abelian group on the m connected components of $U \cap V$, so the image of ∂ is the kernel of the homomorphism from this free abelian group to $\mathbb{Z}$ that maps each component to 1. It is easy to see that this kernel is the free abelian group on $m - 1$ generators, which completes the proof. For example, if the components of $U \cap V$ are numbered $W_1, \ldots, W_m$, and e_i is the class that puts coefficient 1 in front of W_i, -1 in front of W_{i+1}, and 0 in front of the other components, then the classes $e_1, \ldots, e_{m-1}$ form a basis for this kernel.

In particular, this recovers the fact that if U is the complement of a point, then $H_1U = \mathbb{Z}$.

Exercise 10.10. Use Mayer–Vietoris to recover the fact that, if U is the complement of n points in the plane, then H_1U is a free abelian group with n generators. Show that small circles around each of the points gives a basis for H_1U.

10c. Variations and Applications

The assertions of the Mayer–Vietoris story can be put in fancy language (but with no change in content) as follows. Define, for $k=0$ and $k=1$, a homomorphism

$$+ \; : \; H_kU \oplus H_kV \;\to\; H_k(U \cup V)$$

that takes a pair (β, γ) to the sum of the image of β and the image of γ in $H_k(U \cup V)$. In this language, MV(i) says that this homomorphism from $H_0U \oplus H_0V \to H_0(U \cup V)$ is surjective. Similarly, define a homomorphism

$$- \; : \; H_k(U \cap V) \;\to\; H_kU \oplus H_kV$$

that sends a class α to the pair $(\beta, -\gamma)$, where β is the image of α in H_kU and γ is the image of α in H_kV. Assertion MV(ii) says that the image of this homomorphism from $H_0(U \cap V)$ to $H_0U \oplus H_0V$ is the kernel of the homomorphism from $H_0U \oplus H_0V$ to $H_0(U \cup V)$.

Given a sequence $\ldots \to A_{n-1} \to A_n \to A_{n+1} \to \ldots$ of abelian groups and homomorphisms between them, the sequence is called *exact at* A_n if the kernel of the map from A_n to A_{n+1} is equal to the image of the map from A_{n-1} to A_n. The sequence is called *exact* if it is exact at every group in the sequence. For example, the exactness of a sequence $A \to B \to 0$ (at B) means precisely that the map from A to B is surjective, and the exactness of $0 \to A \to B$ (at A) means that the map from A to B is injective. Assertions MV(i)–MV(vi) can be summarized in the

Theorem 10.11 (Mayer–Vietoris Theorem for Homology). *For any open sets U and V of a topological space, the sequence*

$$H_1(U \cap V) \xrightarrow{-} H_1U \oplus H_1V \xrightarrow{+} H_1(U \cup V) \xrightarrow{\partial} H_0(U \cap V) \xrightarrow{-} H_0U \oplus H_0V \xrightarrow{+} H_0(U \cup V) \to 0$$

is exact. If U and V are open subsets of the plane, the sequence

$$0 \to H_1(U \cap V) \xrightarrow{-} H_1U \oplus H_1V \xrightarrow{+} H_1(U \cup V) \xrightarrow{\partial} H_0(U \cap V) \xrightarrow{-} H_0U \oplus H_0V \xrightarrow{+} H_0(U \cup V) \to 0$$

is exact.

Exercise 10.12. Verify that the exactness of the Mayer–Vietoris sequence at each term is indeed equivalent to the assertions MV(i)–MV(vi).

Exercise 10.13. Suppose A and B are disjoint closed subsets in the plane. Use Mayer–Vietoris to construct an isomorphism

$$H_1(\mathbb{R}^2 \setminus (A \cup B)) \cong H_1(\mathbb{R}^2 \setminus A) \oplus H_1(\mathbb{R}^2 \setminus B).$$

Show that the number of connected components of $\mathbb{R}^2 \setminus (A \cup B)$ is one less than the sum of the numbers of connected components of $\mathbb{R}^2 \setminus A$ and of $\mathbb{R}^2 \setminus B$. Generalize to any finite number of disjoint closed subsets.

Exercise 10.14. Let U be an open subset of the plane, and K a compact subset of U. Show that $H_1(U \setminus K) \cong H_1 U \oplus H_1(\mathbb{R}^2 \setminus K)$. If K has n connected components, recover the isomorphism $H_1(U \setminus K) \cong H_1 U \oplus \mathbb{Z}^n$ of Problem 9.9.

Exercise 10.15. If X is a finite graph (see Problem 5.21) with v vertices and e edges, and X has k connected components, show that $H_1 X$ is a free abelian group with $e - v + k$ generators.

Exercise 10.16. If X is a finite graph in the plane with k connected components, show that $H_1(\mathbb{R}^2 \setminus X)$ is a free abelian group with k generators.

Problem 10.17. For any nonempty space X, we considered in §9a the degree homomorphism deg: $H_0 X \to \mathbb{Z}$ that takes the class of a zero cycle $\Sigma n_i P_i$ to the sum Σn_i of the coefficients. The *reduced* 0th homology group $\widetilde{H}_0 X$ of a space X is defined to be the kernel of this map. So X is path-connected if and only if $\widetilde{H}_0 X$ is zero. (a) Show that $\widetilde{H}_0 X$ is a free abelian group with rank one less that the number of path-connected components of X. (b) Show that, if $U \cap V$ is not empty, there is an exact sequence

$$H_1(U \cap V) \xrightarrow{-} H_1 U \oplus H_1 V \xrightarrow{+} H_1(U \cup V)$$
$$\xrightarrow{\partial} \widetilde{H}_0(U \cap V) \xrightarrow{-} \widetilde{H}_0 U \oplus \widetilde{H}_0 V \xrightarrow{+} \widetilde{H}_0(U \cup V) \to 0.$$

This often simplifies computations, since the groups involved are a little smaller.

Problem 10.18. Let U be a connected open set in S^2, and let X be a subset of S^2 homeomorphic to a closed interval such that all of X except for the endpoints is contained in U. Show that $U \setminus X$ is disconnected if and only if the endpoints of X lie in the same connected component of $S^2 \setminus U$.

Problem 10.19. If X is an open set in the plane, and $H_0 X$ and $H_1 X$

have finite ranks, the *Euler characteristic* of X, denoted $\chi(X)$, is defined by

$$\chi(X) \;=\; \operatorname{rank}(H_0X) - \operatorname{rank}(H_1X)\,.$$

Suppose U and V are open sets in the plane, and three of the four open sets U, V, $U \cap V$, and $U \cup V$ have homology of finite ranks. Show that the fourth does also, and that

$$\chi(U \cup V) + \chi(U \cap V) \;=\; \chi(U) + \chi(V)\,.$$

Using the Mayer–Vietoris theorem, homology groups can be used instead of cohomology to prove the Jordan Curve Theorem:

Problem 10.20. With U and V as in the proof of Theorem 5.10, show that the image of the boundary map

$$\partial\colon H_1(U \cup V) = H_1(\mathbb{R}^2 \setminus \{P, Q\}) \;\to\; H_0(U \cap V) = H_0(\mathbb{R}^2 \setminus X)$$

is free with one generator, and deduce that $H_0(\mathbb{R}^2 \setminus X)$ is a free abelian group with two generators, so the complement of X has two connected components. Similarly with U and V as in the proof of Theorem 5.11, show that the image of ∂ is zero, and deduce that $H_0(\mathbb{R}^2 \setminus Y)$ is a free abelian group with one generator, so $\mathbb{R}^2 \setminus Y$ is connected.

The following are some complements to the Jordan Curve Theorem that can be proved by combining Mayer–Vietoris with Corollary 9.4 or Problem 9.9:

Problem 10.21. A compact set K of an open set U is said to *separate* two points if they belong to different connected components of $U \setminus K$. Prove the following version of "Alexander's Lemma": If Y and Z are compact subsets of U, with $Y \cap Z$ connected, and P and Q are two points in $U \setminus (Y \cup Z)$ that are not separated by Y or by Z, then they are not separated by $Y \cup Z$.

Problem 10.22. Let X be a subset of the plane homeomorphic to a circle, P a point on X, and D a disk centered at P. Let A be an open arc of X containing P and contained in D, and let B be the complementary closed arc. Let E be a disk centered at P that doesn't meet B. Show that if two points in E are not separated by X, then they can be connected by a path in $D \setminus X$.

Problem 10.23. Let X be a subset of the plane homeomorphic to a circle, U a connected component of the $\mathbb{R}^2 \setminus X$. Show that, for any

$\varepsilon > 0$, there is a $\delta > 0$ such that any two points of U within distance δ of each other can be connected by a path in the intersection of U with a disk of radius ε. Note that this is false when X is homeomorphic to an interval.

Problem 10.24. Let X be a subset of the plane homeomorphic to a circle, P a point on X, and Q a point not on X. Show that there is a path $\gamma\colon [0, 1] \to \mathbb{R}^2$ with $\gamma(0) = Q$, $\gamma(1) = P$, and $\gamma(t) \notin X$ for $0 \le t < 1$.

Problem 10.25. Suppose Y and Z are compact, connected subsets of an open set U, and $Y \cap Z$ has n connected components. (a) Show that $U \setminus (Y \cup Z)$ has at least n connected components. (b) If $U \setminus Y$ and $U \setminus Z$ are connected, show that $U \setminus (Y \cup Z)$ has exactly n connected components.

10d. Mayer–Vietoris for Cohomology

The Mayer–Vietoris story for cohomology is similar to—in fact, dual to—that for homology. If U' is an open subset of a plane open set U, the restriction of 1-forms from U to U' takes closed forms on U to closed forms on U', and exact forms on U to exact forms on U', so we have a linear map of vector spaces

$$H^1(U) \to H^1(U').$$

There is also a linear map $H^0(U) \to H^0(U')$ that takes a locally constant function on U to its restriction to U'. If U and V are open sets in the plane, we therefore have a diagram

$$\begin{array}{ccccccccc}
 & & H^0U & & & & H^1U & & \\
 & \nearrow & & \searrow & & \nearrow & & \searrow & \\
H^0(U \cup V) & & & & H^0(U \cap V) \xrightarrow{\ \delta\ } H^1(U \cup V) & & & & H^1(U \cap V) \\
 & \searrow & & \nearrow & & \searrow & & \nearrow & \\
 & & H^0V & & & & H^1V & &
\end{array}$$

where the maps in each diagram are determined by restrictions, and δ is the coboundary map defined in Chapter 5. The *Mayer–Vietoris story* for De Rham cohomology, for open sets in the plane, is the combination of six assertions:

MV(i). *An element in $H^0(U \cup V)$ maps to zero in H^0U and in H^0V if and only if it is zero.*

MV(ii). *An element in H^0U and an element in H^0V have the same image in $H^0(U \cap V)$ if and only if they come from some element in $H^0(U \cup V)$.*

MV(iii). *An element in $H^0(U \cap V)$ maps to zero in $H^1(U \cup V)$ if and only if it is the difference of an element coming from H^0U and an element coming from H^0V.*

MV(iv). *An element in $H^1(U \cup V)$ maps to zero in H^1U and H^1V if and only if it is the image by δ of an element in $H^0(U \cap V)$.*

MV(v). *An element in H^1U and an element in H^1V have the same image in $H^1(U \cap V)$ if and only if they come from some element in $H^1(U \cup V)$.*

MV(vi). *Any element in $H^1(U \cap V)$ is the difference of an element coming from H^1U and an element coming from H^1V.*

Of these, MV(i) and MV(ii) are straightforward exercises, and MV(iii) and MV(iv) are Propositions 5.7 and 5.9. We prove the non-trivial assertion in MV(v). Given closed 1-forms α on U and β on V, such that there is a function f on $U \cap V$ with $\alpha - \beta = df$ on $U \cap V$, we must construct a closed 1-form ω on $U \cup V$ that differs from α by an exact form on U, and from β by an exact form on V. That is, we want functions f_1 on U and f_2 on V such that

$$\alpha - df_1 \;=\; \beta - df_2$$

on $U \cap V$. In other words, we want $df_1 - df_2 = df$. So it will be enough to find f_1 and f_2 so that $f_1 - f_2 = f$ on $U \cap V$. The existence of such f_i follows from Lemma 5.5.

The last case, MV(vi), however is not so obvious. We postpone the proof until Chapter 15, where we make rigorous the "duality" between homology and cohomology. With this duality, in fact, all six assertions for cohomology will be seen to follow from the six assertions for homology. □

The cohomology version of Mayer–Vietoris also has its concise expression as an exact sequence. Define linear maps

$$+ \;:\; H^k(U \cup V) \;\to\; H^kU \oplus H^kV$$

by sending a class to the pair consisting of the restrictions of the class to each open set. Define linear maps

$$- \; : \; H^kU \oplus H^kV \; \to \; H^k(U \cap V)$$

by taking a pair (ω_1, ω_2) to the difference $\omega_1|_{U\cap V} - \omega_2|_{U\cap V}$ of its restrictions. As in the case for homology, the assertions MV(i)–MV(vi) are equivalent to the exactness of a sequence:

Theorem 10.26 (Mayer–Vietoris Theorem for Cohomology). *For any open sets U and V in the plane, the sequence*

$$0 \to H^0(U \cup V) \xrightarrow{+} H^0U \oplus H^0V \xrightarrow{-} H^0(U \cap V)$$
$$\xrightarrow{\delta} H^1(U \cup V) \xrightarrow{+} H^1U \oplus H^1V \xrightarrow{-} H^1(U \cap V) \to 0$$

is exact.

It is probably worth remarking that the choice of signs "+" and "−" in both the homology and cohomology versions of Mayer–Vietoris is perfectly arbitrary; those we have chosen will be convenient later.

The following two exercises use Mayer–Vietoris for cohomology, including MV(vi), to strengthen some facts we saw in Chapter 9:

Exercise 10.27. Suppose A and B are disjoint closed subsets in the plane. Use Mayer–Vietoris to construct an isomorphism

$$H^1(\mathbb{R}^2 \setminus (A \cup B)) \; \cong \; H^1(\mathbb{R}^2 \setminus A) \oplus H^1(\mathbb{R}^2 \setminus B).$$

Generalize to the complement of any finite number of disjoint closed subsets.

Exercise 10.28. If K is a compact subset of an open set U, and K has n connected components, construct an isomorphism

$$H^1(U \setminus K) \; \cong \; H^1U \oplus \mathbb{R}^n,$$

where the maps to the second factor are given by periods around the components of K.

Exercise 10.29. If U is not empty, define $\widetilde{H}^0U$ to be the quotient space $H^0U/\mathbb{R}$ of the locally constant functions on U by the subspace of constant functions. This is called the *reduced* 0th cohomology group. So $\dim(\widetilde{H}^0U) = \dim(H^0U) - 1$, and U is connected exactly when $\widetilde{H}^0U$ is zero. Show that, if $U \cap V$ is not empty, there is an exact

sequence

$$0 \to \widetilde{H}^0(U \cup V) \xrightarrow{+} \widetilde{H}^0 U \oplus \widetilde{H}^0 V \xrightarrow{-} \widetilde{H}^0(U \cap V)$$
$$\xrightarrow{\delta} H^1(U \cup V) \xrightarrow{+} H^1 U \oplus H^1 V \xrightarrow{-} H^1(U \cap V) \to 0\,.$$

The result in the following problem was proved by Brouwer:

Problem 10.30. Let K be any compact connected subset of $\mathbb{R}^2$, U a connected component of $\mathbb{R}^2 \setminus K$. Show that the boundary ∂U of U is connected. Is the same true if K is only closed and connected in $\mathbb{R}^2$?

PART VI

COVERING SPACES AND FUNDAMENTAL GROUPS, I

In Chapter 11 we introduce the notion of covering maps, which are generalizations of the polar coordinate mapping, and study their basic properties. Facts about lifting paths and homotopies will generalize what we saw for this special case, which amounted to the basic properties of winding numbers. Many coverings, including the polar coordinate mapping, are examples of G-coverings, arising from an action of a group G on a space, and we emphasize those that arise this way.

We have studied closed paths, and seen that homotopic paths have similar properties. In Chapter 12 we formally introduce the fundamental group, which is the set of homotopy equivalence classes of closed paths starting and ending at a fixed point, the equivalence given by homotopy. In the last section we see how it is related to the first homology group.

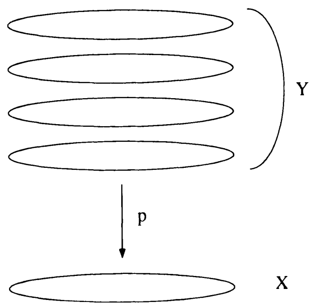
CHAPTER 11

Covering Spaces

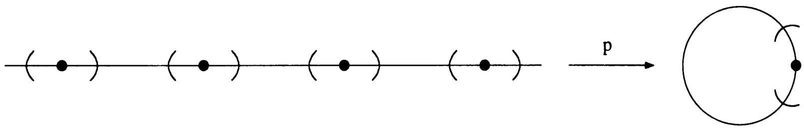
11a. Definitions

If X and Y are topological spaces, a *covering map* is a continuous mapping $p: Y \to X$ with the property that each point of X has an open neighborhood N such that $p^{-1}(N)$ is a disjoint union of open sets, each of which is mapped homeomorphically by p onto N. (If N is connected, these must be the components of $p^{-1}(N)$.) One says that p is *evenly covered* over such N. Such a covering map is called a *covering of* X.

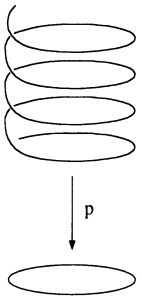
An *isomorphism* between coverings $p: Y \to X$ and $p': Y' \to X$ is a homeomorphism $\varphi: Y \to Y'$ such that $p' \circ \varphi = p$. A covering is called *trivial* if, in the definition, one may take N to be all of X. Equivalently, a covering is trivial if it is isomorphic to the projection of a product $X \times T$ onto X, where T is any set with the discrete topology (all points are closed). So any covering is locally trivial.



The first nontrivial covering is the mapping $p\colon \mathbb{R}\to S^1$ given by $p(\vartheta) = (\cos(2\pi\vartheta), \sin(2\pi\vartheta))$:



or in a vertical picture:



We saw in Chapter 2 that the *polar coordinate* mapping

$$p \ : \ \{(r,\vartheta)\in\mathbb{R}^2\colon r>0\} \ \to \ \mathbb{R}^2\setminus\{(0,0)\}$$

given by $p(r,\vartheta) = (r\cos(\vartheta), r\sin(\vartheta))$, is a covering. Another is the mapping $p_n\colon S^1\to S^1$, for any integer $n\geq 1$, given by

$$(\cos(2\pi\vartheta), \sin(2\pi\vartheta)) \ \mapsto \ (\cos(2\pi n\vartheta), \sin(2\pi n\vartheta)),$$

or, using complex numbers, by $z\to z^n$; this can be visualized by joining the loose ends in the last picture.

Exercise 11.1. Show that the following are covering maps, where $\mathbb{C}^* = \mathbb{C}\setminus\{0\}$ is the group of nonzero complex numbers: (i) the nth power mapping $\mathbb{C}^*\to\mathbb{C}^*$, $z\to z^n$; and (ii) the exponential mapping $\exp\colon \mathbb{C}\to\mathbb{C}^*$.

Exercise 11.2. Let $p\colon Y\to X$ be a covering. If X' is any subspace of X, verify that the restriction $Y' = p^{-1}(X')\to X'$ is a covering.

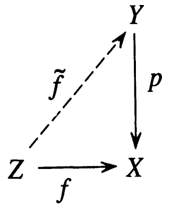
Exercise 11.3. If $p\colon Y\to X$ is a covering of an open subset X in the plane, show that Y can be given the structure of a differentiable surface in such a way that p is a local diffeomorphism. The components V of $p^{-1}(N)$, for N a connected open set in X over which the covering

is trivial, together with the homeomorphisms of N with V given by the inverse of p, give charts covering Y. Similarly, any covering of any manifold has a natural manifold structure.

Exercise 11.4. Show that if X is connected, all fibers of a covering $Y \to X$ have the same cardinality.

When each $p^{-1}(x)$ has cardinality a finite number n, the covering is called an *n-sheeted* covering. It should be emphasized, however, that one cannot distinguish n different "sheets," unless the covering is trivial.

If $p: Y \to X$ is a covering, and $f: Z \to X$ is a continuous mapping, a continuous mapping $\tilde{f}: Z \to Y$ such that $p \circ \tilde{f} = f$ is called a *lifting* of f.



In the next section we will discuss the question of whether such liftings exist. The following lemma discusses uniqueness: if Z is connected, a lifting is determined by where it maps any one point.

Lemma 11.5. *Let $p: Y \to X$ be a covering, and let Z be a connected topological space. Suppose $\tilde{f}_1$ and $\tilde{f}_2$ are continuous mappings from Z to Y such that $p \circ \tilde{f}_1 = p \circ \tilde{f}_2$. If $\tilde{f}_1(z) = \tilde{f}_2(z)$ for one point z in Z, then $\tilde{f}_1 = \tilde{f}_2$.*

Proof. It suffices to show that the set in Z where the mappings agree is open, and its complement where they disagree is also open. If w is in the set where they agree, take a neighborhood N of $p \circ \tilde{f}_i(w)$ that is evenly covered by p. Let $p^{-1}(N)$ be a disjoint union of open sets N_α, with each N_α, mapped homeomorphically to N by p. By continuity, $\tilde{f}_1$ and $\tilde{f}_2$ must map a neighborhood V of w into the same N_α, and since $p \circ \tilde{f}_1 = p \circ \tilde{f}_2$, $\tilde{f}_1$ and $\tilde{f}_2$ must agree on V. Similarly, if w is in the set where the mappings do not agree, $\tilde{f}_1$ and $\tilde{f}_2$ must map a neighborhood V of w into two different (and hence disjoint) N_α's, so they disagree on V. □

11b. Lifting Paths and Homotopies

The work we did in studying the winding number will be abstracted and formalized in the following propositions. In §2a we saw that defining an angle function amounted to lifting a closed path γ in $\mathbb{R}^2 \setminus \{0\}$ to a path $\tilde{\gamma}$ in the right half plane so that the composite of $\tilde{\gamma}$ with the polar coordinate covering p is the given path γ; the difference in second coordinates of $\tilde{\gamma}$ from start to end is then the change in angle along γ. We will now see that this is a general property of covering spaces.

Proposition 11.6 (Path Lifting). *Let $p\colon Y \to X$ be a covering, and let $\gamma\colon [a,b] \to X$ be a continuous path in X. Let y be a point of Y with $p(y) = \gamma(a)$. Then there is a unique continuous path $\tilde{\gamma}\colon [a,b] \to Y$ such that $\tilde{\gamma}(a) = y$ and $p \circ \tilde{\gamma}(t) = \gamma(t)$ for all t in the interval $[a,b]$.*

Proof. The uniqueness comes from Lemma 11.5. When the covering is trivial, the proposition is obvious: there is a unique component of Y that contains y and maps homeomorphically to the component of X that contains $\gamma([a,b])$, and one must simply lift the path to that component using the inverse homeomorphism. For the general case, we apply the Lebesgue lemma (Appendix A) to the open sets $\gamma^{-1}(N)$, where N varies over open sets in X that are evenly covered by p. This gives a subdivision $a = t_0 \leq \ . \ . \ . \ \leq t_n = b$ such that each $\gamma([t_{i-1}, t_i])$ is contained in some open set that is evenly covered by p. By the trivial case, there is a lifting of the restriction of γ to $[t_0, t_1]$, giving a path in Y from y to some point y_1. Similarly, there is a lifting of the restriction of γ to $[t_1, t_2]$ that starts at y_1; and one proceeds in n steps until one has lifted the whole path. □

It follows in particular that the final point $\tilde{\gamma}(b)$ of the lifting is determined by γ and by the initial point y. We denote this point by $y * \gamma$, so

$$y * \gamma \ = \ \tilde{\gamma}(b).$$

If this is applied to the polar coordinate covering, and $y = (r_0, \vartheta_0)$ and $y * \gamma = (r_1, \vartheta_1)$ are the initial and final points of the lifting of a path γ, then the difference $\vartheta_1 - \vartheta_0$ is the total change of angle along γ.

Exercise 11.7. Verify this last statement.

One consequence of the path-lifting proposition is the fact that any covering of an interval must be trivial.

The fact that homotopic paths in $\mathbb{R}^2 \setminus \{0\}$ with the same endpoints have the same total change of angle around zero is equivalent to the fact that their liftings are homotopic. This too is a general fact about coverings.

Proposition 11.8 (Homotopy Lifting). *Let $p: Y \to X$ be a covering, and let H be a homotopy of paths in X, i.e., $H: [a,b] \times [0,1] \to X$ is a continuous mapping. Let $\gamma_0(t) = H(t,0)$, $a \le t \le b$, be the initial path. Suppose $\tilde{\gamma}_0$ is a lifting of γ_0. Then there is a unique lifting $\tilde{H}$ of H whose initial path is $\tilde{\gamma}_0$, i.e., $\tilde{H}: [a,b] \times [0,1] \to Y$ is continuous, with $p \circ \tilde{H} = H$ and $\tilde{H}(t,0) = \tilde{\gamma}_0(t)$, $a \le t \le b$.*

Proof. The proof is very much the same as for Proposition 11.6. First apply the Lebesgue lemma to know that there are subdivisions $a = t_0 < t_1 < \ldots < t_n = b$ and $0 = s_0 < s_1 < \ldots < s_m = 1$ so that if $R_{i,j}$ is the rectangle $[t_{i-1}, t_i] \times [s_{j-1}, s_j]$, then each $H(R_{i,j})$ is contained in an evenly covered open set. Then the lifting $\tilde{H}$ is constructed over each piece $R_{i,j}$, say first working across the bottom row, lifting the restriction of H to $R_{1,1}, R_{2,1}, \ldots, R_{n,1}$, then doing the same for the next row $R_{1,2}, R_{2,2}, \ldots, R_{n,2}$, and so on until the entire rectangle has been covered. □

If H is a homotopy of paths from x to x' in X, i.e., $H(a,s) = x$ and $H(b,s) = x'$ for all $0 \le s \le 1$, and if $\tilde{\gamma}_0$ is a path from y to y', then the lifted homotopy $\tilde{H}$ is a homotopy of paths from y to y'. (The fact that $\tilde{H}$ is constant on the sides of the rectangle is guaranteed by the uniqueness of the lifting of the restriction of H to these sides.) In particular, if H is a homotopy from γ_0 to γ_1, then

$$y * \gamma_0 \;=\; y * \gamma_1 .$$

For the polar coordinate mapping, this is the preceding assertion about homotopic paths having the same change of angle.

Exercise 11.9. Use the homotopy lifting proposition to show that homotopic closed paths in $\mathbb{R}^2 \setminus \{0\}$ have the same winding number. (The paths in the homotopy are assumed to be closed, but the endpoints can vary during the homotopy.)

Exercise 11.10. Use the homotopy lifting proposition to prove that any covering of a rectangle (closed or open) must be trivial. Deduce the same for any space homeomorphic to a rectangle, such as a disk.

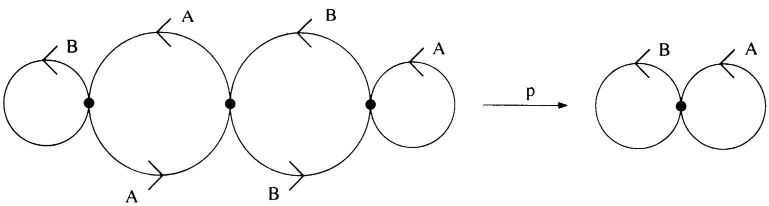
Exercise 11.11. Suppose $p: Y \to X$ is a covering map, with X a locally

connected space (any neighborhood of a point contains a connected neighborhood of the point). (a) Show that X is a union of connected open sets N such that each connected component of $p^{-1}(N)$ is mapped homeomorphically by p onto N. (b) Let Y' be a connected component of Y. Show that image $X' = p(Y')$ is a connected component of X, and the restriction $Y' \to X'$ is a covering map.

Exercise 11.12. Suppose $p: Y \to X$ and $p': Y' \to X$ are covering maps, and $\varphi: Y \to Y'$ is a continuous map such that $p' \circ \varphi = p$. Suppose X, Y, and Y' are connected, and X is locally connected. Show that φ is a covering map.

Exercise 11.13. If $p: S' \to S$ is an n-sheeted covering, and S is a compact surface, show that S' is also a compact surface. Show, using vector fields and/or triangulations, that the Euler characteristic of S' is n times the Euler characteristic of S. If S and S' are spheres with g and g' handles, show that $g' = ng - n + 1$.

Exercise 11.14. Let X be the space that consist of two circles A and B joined at a point P. Let Y be a space that is two circles and four half-circles, joined as shown, and let $p: Y \to X$ be mapping that takes each piece of Y to the correspondingly labeled piece of X as indicated:



Show that p is a three-sheeted covering. Let γ be the path in X, starting at P, that goes first around the circle A counterclockwise, then around B counterclockwise, then around A clockwise, then around B clockwise. Find the three liftings of γ. Deduce that γ is not homotopic to the constant path at P. Use this to solve Problem 9.14.

11c. G-Coverings

Many important covering spaces arise from the action of a group G on a space Y, with X the space of orbits. Recall that an *action* of a

group G on Y (on the left) is a mapping $G \times Y \to Y$, $(g, y) \mapsto g \cdot y$, satisfying:

(1) $g \cdot (h \cdot y) = (g \cdot h) \cdot y$ for all g and h in G and y in Y;
(2) $e \cdot y = y$ for all y in Y, where e is the identity in G; and
(3) the mapping $y \mapsto g \cdot y$ is a homeomorphism of Y for all g in G.

In other words, G defines a group of homeomorphisms of Y. Two points y and y' are in the same *orbit* if there is an element g in G that maps one to the other: $y' = g \cdot y$. Since G is a group, this is an equivalence relation. Let $X = Y/G$ be the set of orbits, or equivalence classes. There is a *projection* p: $Y \to X$ that maps a point to the orbit containing it. The space X is equipped with the quotient topology (i.e., a set U in X is defined to be open when $p^{-1}(U)$ is open in Y).

Exercise 11.15. (a) The group $\mathbb{Z}$ acts on $\mathbb{R}$ by translation: $n \cdot r = r + n$. Show that the quotient $\mathbb{R} \to \mathbb{R}/\mathbb{Z}$ can be identified with the covering map from $\mathbb{R}$ to S^1 described in the first section.

(b) Show that the polar coordinate covering map is the quotient of the right half plane by a $\mathbb{Z}$-action.

(c) The group $G = \mu_n$ of nth roots of unity (which is a cyclic group of order n) in $\mathbb{C}$ acts by multiplication on S^1, regarded as the complex numbers of absolute value 1. Compare the quotient map with the covering p_n: $S^1 \to S^1$ described in §11a.

For another example, the group with two elements acts on the n-sphere with the nontrivial element taking a point to its antipodal point; i.e., $G = \{\pm 1\} \cong \mathbb{Z}/2\mathbb{Z}$ acts on S^n, with $\pm 1 \cdot P = \pm P$. The quotient space is the real projective space $\mathbb{RP}^n$, and the quotient mapping $S^n \to \mathbb{RP}^n$ is a two-sheeted covering.

We say that G acts *evenly*[6] if any point in Y has a neighborhood V such that $g \cdot V$ and $h \cdot V$ are disjoint for any distinct elements g and h in G.

Exercise 11.16. The group μ_n of nth roots of unity acts on $\mathbb{C}$ by

[6] The standard terminology for the notion we are calling "even" is the mouthful "properly discontinuous." The word "discontinuous" does *not* mean that anything is not continuous or otherwise badly behaved; it means that the orbits are discrete subsets of Y. The word "properly" refers to the fact that every compact set only meets finitely many of its translates. If this were not bad enough, the use of "properly discontinuously" in the literature is inconsistent, in that often it is only required that each point have a neighborhood V such that at most finitely many translates $g \cdot V$ of V can intersect. In this case the word "freely" is added, so our "evenly" is then "freely and properly discontinuously."

multiplication. Show that the action is not even, but that the action on the open subset $\mathbb{C}^* = \mathbb{C} \setminus \{0\}$ is even.

Lemma 11.17. *If a group G acts evenly on Y, then the projection $p: Y \to Y/G$ is a covering map.*

Proof. The map p is continuous, and open since for V open in Y, the set $p^{-1}(p(V))$ is the union of open sets $g \cdot V$, $g \in G$. If V is taken as in the definition of an even action, then this union is a disjoint union. It is enough to prove that for such V, the mapping from each $g \cdot V$ to $p(V)$ induced by p is a bijection, for then it follows that p is evenly covered over $p(V)$. This is a straightforward verification: since $p(g \cdot y) = p(y)$ for y in V, it is surjective; if $p(g \cdot y_1) = p(g \cdot y_2)$, there is some h in G with $h \cdot g \cdot y_1 = g \cdot y_2$, and the fact that the action is even implies that h is the identity. □

A covering $p: Y \to X$ is called a *G-covering* if it arises in this way from an even action of G on Y. An *isomorphism* of G-coverings is an isomorphism of coverings that commutes with the action of G; i.e., an isomorphism of the G-covering $p: Y \to X$ with the G-covering $p': Y' \to X$ is a homeomorphism $\varphi: Y \to Y'$ such that $p' \circ \varphi = p$ and $\varphi(g \cdot y) = g \cdot \varphi(y)$ for g in G and y in Y. The *trivial* G-covering of X is the product $X \times G \to X$, where G acts by (left) multiplication on the second factor.

Lemma 11.18. *Any G-covering is locally trivial as a G-covering, i.e., if $p: Y \to X$ is a G-covering, then any point in X has a neighborhood N such that the G-covering $p^{-1}(N) \to N$ is isomorphic to the trivial G-covering $N \times G \to N$.*

Proof. In fact, if $N = p(V)$ as in the proof of Lemma 11.17, such a local trivialization is given by

$$p^{-1}(N) \ni g \cdot v \quad \mapsto \quad p(v) \times g \in N \times G. \qquad \square$$

Exercise 11.19. Verify that the actions described in Exercise 11.15 are all even. Show that the action of $\mathbb{Z}^n$ on $\mathbb{R}^n$ by translation is even. Identify the quotient $\mathbb{R}^n/\mathbb{Z}^n$ with the n-dimensional torus $(S^1)^n = S^1 \times \ldots \times S^1$.

Exercise 11.20. Show that any two-sheeted covering has a unique structure of G-covering, where $G = \mathbb{Z}/2\mathbb{Z}$ is the group of order two.

Exercise 11.21. Show that the three-sheeted covering of Exercise 11.14 is not a G-covering.

Exercise 11.22. A *section* of a covering $p: Y \to X$ is a continuous mapping $s: X \to Y$ such that $p \circ s$ is the identity mapping of X. Show that if a G-covering has a section, then the covering is a trivial G-covering.

Exercise 11.23. If $p: Y \to X = Y/G$ is a G-covering that is trivial as a covering, show that it is isomorphic to the trivial G-covering.

Exercise 11.24. Let $p: Y \to X = Y/G$ be a G-covering, and let φ_1 and φ_2 be isomorphisms of G-coverings from Y to Y. If X is connected, and φ_1 and φ_2 agree at one point of Y, show that $\varphi_1 = \varphi_2$.

Exercise 11.25. Let G be the subgroup of the group of homeomorphisms of the plane to itself generated by the translation $(x, y) \mapsto (x+1, y)$ and by the mapping $(x, y) \mapsto (-x, y+1)$. Show that this action of G on $\mathbb{R}^2$ is even, and identify the quotient $\mathbb{R}^2/G$ with the Klein bottle.

Exercise 11.26. Let G be the subgroup of the group of homeomorphisms of the plane to itself generated by the translation $(x, y) \mapsto (x+1, -y)$. Show that this action is even, and identify the quotient with a Moebius band.

Exercise 11.27. If G acts evenly on a space Y, and H is a subgroup of G, show that H also acts evenly. Show that the natural map from Y/H to Y/G is a covering mapping. If n is the index of H in G, this is an n-sheeted covering. Carry this out when G is the group of transformations of the plane from Exercise 11.25, and H is the subgroup generated by the two homeomorphisms $(x, y) \mapsto (x+1, y)$ and $(x, y) \mapsto (x, y+2)$. Identify $\mathbb{R}^2/H$ with a torus, and describe the resulting two-sheeted covering of the Klein bottle.

Exercise 11.28. If a finite group G acts on a Hausdorff space Y, and there are no fixed points (i.e., no y is fixed by any g in G except the identity element), show that the action is even.

Exercise 11.29. If G acts evenly as a group of diffeomorphisms of a differentiable manifold Y, show how to give Y/G the structure of a differentiable manifold in such a way that the projection $Y \to Y/G$ is a local diffeomorphism.

Exercise 11.30. Let $G = \mu_n$ be the group of nth roots of unity, as in Exercise 11.15. The odd-dimensional sphere S^{2m-1} can be realized as

$$S^{2m-1} = \{(z_1, \ldots, z_m) \in \mathbb{C}^m : |z_1|^2 + \ldots + |z_m|^2 = 1\}.$$

The group G acts on S^{2m-1} by $\zeta \cdot (z_1, \ldots, z_m) = (\zeta z_1, \ldots, \zeta z_m)$. Show that this action is even. When n is prime, the quotient space S^{2m-1}/μ_n is a manifold called a *Lens space*.

Exercise 11.31. Suppose $\mathfrak{G}$ is a topological group, i.e., a topological space which is also a group such that the multiplication and inverse maps are continuous. Suppose G is a discrete subgroup of $\mathfrak{G}$, i.e., there is a neighborhood N of the identity e in $\mathfrak{G}$ such that $N \cap G = \{e\}$. Show that left multiplication by G on $\mathfrak{G}$ is an even action. This makes $\mathfrak{G}$ a G-covering of the space $G \backslash \mathfrak{G}$ of right cosets of G in $\mathfrak{G}$. Many of our coverings have this form, e.g., $\mathbb{R}^n \to (S^1)^n$ is the quotient by the subgroup $\mathbb{Z}^n$; exp: $\mathbb{C} \to \mathbb{C}^*$ is the quotient of $\mathbb{C}$ by $2\pi i\mathbb{Z}$.

Exercise 11.32. Let $Y = \mathbb{R}^n \setminus \{0\}$, let r be any real number but 0, 1, or -1. Let $G = \mathbb{Z}$ act on Y by $m \cdot v = r^m v$ for $m \in \mathbb{Z}$, $v \in Y$. Show that the action is even, and show that Y/G is homeomorphic to the product $S^1 \times S^{n-1}$.

Problem 11.33 (For those who know about quaternions). The three-sphere S^3 can be identified with the set of unit quaternions:

$$S^3 = \{r + x\mathbf{i} + y\mathbf{j} + z\mathbf{k}: r^2 + x^2 + y^2 + z^2 = 1\},$$

which makes it a topological group. Identify $\mathbb{R}^3$ with the set of pure quaternions $\{x\mathbf{i} + y\mathbf{j} + z\mathbf{k}\}$. For $\mathbf{q} \in S^3$, the mapping $\mathbf{v} \mapsto \mathbf{q} \cdot \mathbf{v} \cdot \mathbf{q}^{-1}$ defines an orthogonal transformation of $\mathbb{R}^3$ with determinant 1, i.e., an element of SO(3). Show that the resulting map $S^3 \to \mathrm{SO}(3)$ is a surjective homomorphism of groups, with kernel $\{\pm 1\}$. Deduce that this is two-sheeted covering. (This is the *spin* group Spin(3), which, with its generalizations $\mathrm{Spin}(n) \to \mathrm{SO}(n)$ for all $n \geq 3$, appear frequently in modern mathematics and physics.)

Problem 11.34. Suppose G is a subgroup of the group of distance-preserving transformations of the plane $\mathbb{R}^2$, that satisfies a *uniformity* condition: there is a $d > 0$ such that for all points P in the plane and all g in G other than the identity, $\|g \cdot P - P\| \geq d$.

(a) Show that the action of G on the plane $\mathbb{R}^2$ is even. Let p: $\mathbb{R}^2 \to \mathbb{R}^2/G = X$ be the resulting G-covering. Define a distance function on X by

$$\mathrm{dist}(Q_1, Q_2) = \mathrm{Min}\{\|P_1 - P_2\|: p(P_1) = Q_1 \text{ and } p(P_2) = Q_2\}.$$

(b) Show that this distance function defines a metric on X.

(c) Prove the analogous result for uniform actions on $\mathbb{R}^3$ (or on any

locally compact metric space). This puts a geometric structure on the quotient space X which is locally Euclidean. The reader is invited to look at the resulting "locally Euclidean geometry" on X for some of the groups we have seen. For a general discussion, with a list of possibilities for $\mathbb{R}^2$ and $\mathbb{R}^3$, see Nikulin and Shafarevich (1987).

11d. Covering Transformations

For any covering $p: Y \to X$ there is a group $\mathrm{Aut}(Y/X)$ of *covering transformations*, or *deck transformations*:

$$\mathrm{Aut}(Y/X) = \{\varphi: Y \to Y : \varphi \text{ is a homeomorphism and } p \circ \varphi = p\}.$$

This is a group by composition of mappings, and it acts on Y in the sense of the preceding section; it is called the *automorphism group* of the covering.

Exercise 11.35. If $Y \to X$ is a trivial n-sheeted covering, and X is connected, show that $\mathrm{Aut}(Y/X)$ is isomorphic to the symmetric group on n letters.

Exercise 11.36. If $p: Y \to X$ is the covering of Exercise 11.14, show that $\mathrm{Aut}(Y/X)$ contains only the identity element.

If the covering is a G-covering, there is a canonical homomorphism from G to $\mathrm{Aut}(Y/X)$ that takes g to the homeomorphism $y \mapsto g \cdot y$. By the definition of even action, this homomorphism is injective. It need not be surjective; for example, it is not surjective if the covering is a trivial G-covering, see Exercise 11.35.

Proposition 11.37. *If $p: Y \to X$ is a G-covering, and Y is connected, then the canonical homomorphism $G \to Aut(Y/X)$ is an isomorphism.*

Proof. Fix any point y in Y. Given φ in $\mathrm{Aut}(Y/X)$, since y and $\varphi(y)$ lie in the same orbit, there is a g in G such that $g \cdot y = \varphi(y)$. The covering transformation determined by g and φ both take y to $\varphi(y)$, so by Lemma 11.5 they coincide. □

For a G-covering, G acts transitively on each fiber $p^{-1}(x)$ of p; that is, for y and y' in a fiber, there is a g in G with $g \cdot y = y'$. In addition, this action is faithful: the element g taking y to y' is unique. The following proposition gives a converse to this:

Proposition 11.38. *Let $p: Y \to X$ be a covering, with Y connected and X locally connected. Then $Aut(Y/X)$ acts evenly on Y. If $Aut(Y/X)$ acts transitively on a fiber of p, then the covering is a G-covering, with $G = Aut(Y/X)$.*

Proof. We show first that the action is even. For y in Y, let N be a neighborhood of $p(y)$ evenly covered by p, and let V be the neighborhood of y mapping homeomorphically to N by p. If φ and φ' are distinct covering transformations, then $\varphi(V)$ and $\varphi'(V)$ must be disjoint, for if not, then $\varphi^{-1} \circ \varphi'$ has a fixed point in V, and Lemma 11.5 implies that $\varphi^{-1} \circ \varphi'$ is the identity.

Let $G = \text{Aut}(Y/X)$. The covering p factors into the composite of the projection $Y \to Y/G$ followed by a mapping $\bar{p}: Y/G \to X$. It follows easily from the definitions that this $\bar{p}$ is a covering mapping; indeed, any open connected set of X that is evenly covered by p is evenly covered by $\bar{p}$. If x is a point in X such that G acts transitively on $p^{-1}(x)$, then the fiber of $\bar{p}$ over x has only one point. By Exercise 11.4, all fibers of $\bar{p}$ have one point, so $\bar{p}$ is a homeomorphism. It follows that p is a G-covering. □

Problem 11.39. Let $p: Y \to X$ be a covering, with Y connected and X path-connected. Let $x \in X$, and set $S = p^{-1}(x)$. Any automorphism of the covering restricts to a permutation of S, and the automorphism is determined by this restriction, so $\text{Aut}(Y/X) \subset \text{Aut}(S)$. Show that

$$\text{Aut}(Y/X) \cong \{\varphi \in \text{Aut}(S) : \varphi(z * \sigma) = \varphi(z) * \sigma \quad \text{for all} \quad z \in S \text{ and all closed paths } \sigma \text{ at } x\}.$$

CHAPTER 12

The Fundamental Group

12a. Definitions and Basic Properties

The aim of this section is to make the homotopy equivalence classes of paths that start and end at a fixed point in a space into a group. We will see later how this group tells us about the covering spaces of X, as well as determining the first homology and cohomology groups (when X is an open set in the plane). In this chapter it will be convenient again to have all paths defined on the same interval. So a *path* in a topological space X will be a continuous map $\gamma\colon [0,1] \to X$. We say that γ is a path *from* the point $x = \gamma(0)$ *to* the point $x' = \gamma(1)$. In this chapter a *homotopy* of paths will always fix the endpoints x and x', i.e., $H(0,s) = x$ and $H(1,s) = x'$ for all $0 \le s \le 1$.

If σ is a path from a point x to a point x', and τ is a path from x' to a point x'', there is a *product* path, denoted $\sigma \cdot \tau$, which is a path from x to x''. It is the path that first traverses σ, then τ, but it must do so at double speed to complete the trip in the same unit time:

$$\sigma \cdot \tau(t) \;=\; \begin{cases} \sigma(2t), & 0 \le t \le 1/2, \\ \tau(2t-1), & 1/2 \le t \le 1. \end{cases}$$

If σ is a path from x to x', there is an *inverse* path σ^{-1} from x' to x:

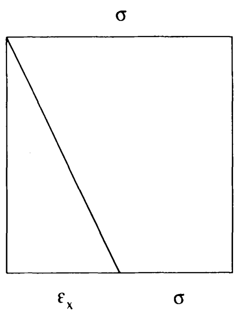
$$\sigma^{-1}(t) \;=\; \sigma(1-t), \qquad 0 \le t \le 1.$$

For any point x, let ε_x be the *constant* path at x:

$$\varepsilon_x(t) \;=\; x, \qquad 0 \le t \le 1.$$

Exercise 12.1. Show that, for paths from a point x to a point x', the relation of being homotopic is an equivalence relation.

Next we verify that, up to homotopy, these operations satisfy (where defined!) the group axioms. For example, if σ is a path from x to x', there is a homotopy from the path $\varepsilon_x \cdot \sigma$ to σ. This is done by adjusting the time of waiting at x, as indicated in the diagram:



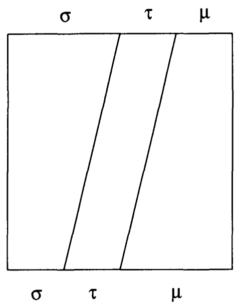
We write down the corresponding homotopy, which is constant on the vertical and slanted lines drawn; on horizontal lines, it is the same as indicated on the top and bottom, but adjusted proportionally:

$$H(t,s) = \begin{cases} x, & 0 \le t \le {}^1\!/_2(1-s), \\ \sigma\left(\dfrac{t - {}^1\!/_2(1-s)}{1 - {}^1\!/_2(1-s)}\right), & {}^1\!/_2(1-s) \le t \le 1. \end{cases}$$

Exercise 12.2. Construct a homotopy from $\sigma \cdot \varepsilon_{x'}$ to σ.

In these cases, and those that follow, it is not hard to write down explicit formulas for the homotopy, using a little plane geometry to map rectangles and triangles onto each other, to interpolate between the values we want on boundary lines. It may be worth pointing out that one could also appeal to general results about maps between convex sets that guarantee the existence of such maps; the formulas themselves are not important.

If σ is a path from x to x', τ a path from x' to x'', and μ is a path from x'' to x''', there is a homotopy from $(\sigma \cdot \tau) \cdot \mu$ to $\sigma \cdot (\tau \cdot \mu)$ constructed similarly from the diagram

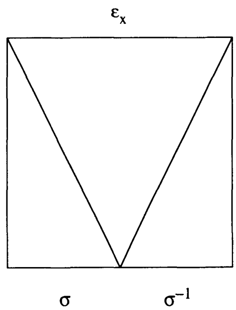


Exercise 12.3. (a) Write down the homotopy indicated by this diagram. (b) Define the path $\sigma \cdot \tau \cdot \mu$ by the formula

$$\sigma \cdot \tau \cdot \mu(t) = \begin{cases} \sigma(3t), & 0 \leq t \leq 1/3, \\ \tau(3t-1), & 1/3 \leq t \leq 2/3, \\ \mu(3t-2), & 2/3 \leq t \leq 1. \end{cases}$$

Show that $\sigma \cdot \tau \cdot \mu$ is homotopic to the paths $(\sigma \cdot \tau) \cdot \mu$ and $\sigma \cdot (\tau \cdot \mu)$.

There is also a homotopy from the path $\sigma \cdot \sigma^{-1}$ to the constant path ε_x. This family of paths does part of the trip specified by σ, rests at the point it has reached, then returns. This is indicated on the following diagram, but note that this time the function is not constant on the diagonal lines:



The homotopy is

$$H(t,s) = \begin{cases} \sigma(2t), & 0 \leq t \leq 1/2(1-s), \\ \sigma(1-s), & 1/2(1-s) \leq t \leq 1/2(1+s), \\ \sigma(2-2t), & 1/2(1+s) \leq t \leq 1. \end{cases}$$

We need one more homotopy: if two paths σ and σ' are homotopic by a homotopy H_1, and τ and τ' are homotopic by a homotopy H_2, then $\sigma \cdot \tau$ and $\sigma' \cdot \tau'$ are homotopic, by the homotopy

$$H(t,s) = \begin{cases} H_1(2t,s), & 0 \le t \le 1/2, \\ H_2(2t-1,s), & 1/2 \le t \le 1. \end{cases}$$

Now, for a point x in X, a *loop at* x is a path that starts and ends at x. Define the *fundamental group of* X *with base point* x, denoted $\pi_1(X,x)$, to be the set of equivalence classes of loops at x, where the equivalence is by homotopy (see Exercise 12.1). We write $[\gamma]$ for the class of the loop γ. The *identity* is the class $e = [\varepsilon_x]$ of the constant path ε_x. Define a product by $[\sigma] \cdot [\tau] = [\sigma \cdot \tau]$. The last displayed homotopy implies that this product is well defined on the equivalence classes. The others imply the equations

$$e \cdot [\sigma] = [\sigma], \qquad [\sigma] \cdot e = e, \qquad [\sigma] \cdot [\sigma^{-1}] = e,$$
$$([\sigma] \cdot [\tau]) \cdot [\mu] = [\sigma] \cdot ([\tau] \cdot [\mu]),$$

in $\pi_1(X,x)$. Since $(\sigma^{-1})^{-1} = \sigma$, the equation $[\sigma^{-1}] \cdot [\sigma] = e$ follows. So this product makes $\pi_1(X,x)$ into a group. From Exercise 12.3(b) we deduce that the product $([\sigma] \cdot [\tau]) \cdot [\mu]$ is also equal to $[\sigma \cdot \tau \cdot \mu]$.

It should be emphasized that endpoints must be fixed during the homotopies discussed here. Otherwise any loop γ would be homotopic to a constant loop, by the homotopy $H(t,s) = \gamma((1-s)t)$.

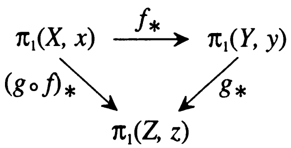
If $f\colon X \to Y$ is a continuous function, and $f(x) = y$, then f determines a homomorphism of groups

$$f_*\colon \pi_1(X,x) \to \pi_1(Y,y),$$

that takes $[\sigma]$ to $[f \circ \sigma]$.

Exercise 12.4. Verify that this is well defined and a group homomorphism.

The fundamental group is a "covariant functor, on the category of pointed spaces." This means that, if we also have a map $g\colon Y \to Z$ with $g(y) = z$, so that we also have $g_*\colon \pi_1(Y,y) \to \pi_1(Z,z)$, then $(g \circ f)_* = g_* \circ f_*$, i.e., the diagram



commutes. In addition, if f is the identity, then $f_*\colon \pi_1(X,x) \to \pi_1(X,x)$ is the identity mapping.

Some of the simple applications of fundamental groups use nothing more than this functoriality. For example, if we know that $\pi_1(X,x)$ is a "complicated" group, and $\pi_1(Y,y)$ is a "simple" group, so that there are no homomorphisms of groups such that the composite $\pi_1(X,x) \to \pi_1(Y,x) \to \pi_1(X,x)$ is the identity, then there can be no continuous maps $f\colon X \to Y$ and $g\colon Y \to X$ such that $g \circ f$ is the identity map on X. For example, if $\pi_1(X,x)$ is not the trivial group, and $\pi_1(Y,y)$ is trivial, then X cannot be embedded in Y as a retract.

Exercise 12.5. Show that $\pi_1(D^2,x) = \{e\}$ for any x in the disk, and that the winding number determines an isomorphism of $\pi_1(S^1,(1,0))$ with $\mathbb{Z}$. Deduce that S^1 is not a retract of D^2.

Exercise 12.6. Show that if $X \subset \mathbb{R}^n$ is a subspace that is starshaped about the point x, then $\pi_1(X,x) = \{e\}$.

Although the definition of fundamental group depends on the choice of base point, one gets the same group (up to isomorphism) if one chooses another base point, at least if X is path-connected, or if the two points can be connected by a path. Suppose τ is a path from x to x'. Define a map

$$\tau_\#\colon \ \pi_1(X,x) \ \to \ \pi_1(X,x')$$

by $[\gamma] \mapsto [\tau^{-1} \cdot (\gamma \cdot \tau)] = [(\tau^{-1} \cdot \gamma) \cdot \tau]$. As before, this is well defined, and is a homomorphism of groups. It is an isomorphism, since $(\tau^{-1})_\#$ gives the inverse homomorphism from $\pi_1(X,x')$ to $\pi_1(X,x)$.

Exercise 12.7. (a) Verify these assertions. (b) If τ' is another path from x to x', show that

$$(\tau')_\#[\gamma] \ = \ [\rho]^{-1} \cdot (\tau_\#[\gamma]) \cdot [\rho],$$

where ρ is the loop $\tau^{-1} \cdot \tau'$. In particular, if τ and τ' are homotopic paths from x to x', they determine the same isomorphism on fundamental groups. In general, the displayed equation means that the isomorphism from $\pi_1(X,x)$ to $\pi_1(X,x')$ depends on the choice of path from x to x' only up to inner automorphism.

For this reason, if X is a path-connected space, one often speaks of "the fundamental group" of X, without referring to a base point. This will cause no confusion as long as we are only interested in the group up to isomorphism.

The fundamental group of a ball D^n or of $\mathbb{R}^n$ is trivial, say by Exercise 12.6. We have seen that the circle has an infinite cyclic fundamental group. Let's look next at the n-sphere S^n, for $n \geq 2$. This is almost as simple as it seems. The complement of any point in S^n is homeomorphic to $\mathbb{R}^n$, so any loop that misses any point is homotopic to the constant loop. Although it is possible for a continuous loop to map onto the n-sphere, for any such n, this is not a serious obstruction. We will see a more general reason later, but for now it can be seen directly:

Exercise 12.8. Show that any path in S^n, $n \geq 2$, is homotopic to a path whose image contains no neighborhood of any of its points. Deduce that the fundamental group of S^n is trivial if $n \geq 2$.

Exercise 12.9. Show that the fundamental group of a Cartesian product is the product of fundamental groups of the spaces:

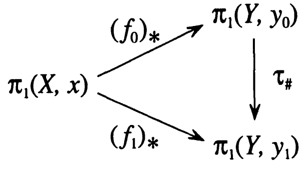
$$\pi_1(X \times Y, x \times y) \cong \pi_1(X, x) \times \pi_1(Y, y).$$

Problem 12.10: If $\mathfrak{G}$ is a topological group, show that $\pi_1(\mathfrak{G}, e)$ is commutative.

12b. Homotopy

Two continuous maps $f_0: X \to Y$ and $f_1: X \to Y$ are *homotopic* if there is a continuous mapping $H: X \times [0, 1] \to Y$ such that $H(z, 0) = f_0(z)$ and $H(z, 1) = f_1(z)$ for all points z in X; H is a *homotopy from f_0 to f_1*. We want to say that homotopic maps determined the same homomorphism on fundamental groups, but to make this precise we must keep track of base points. Let x be a base point in X, and let $y_0 = f_0(x)$ and $y_1 = f_1(x)$. The mapping $\tau(t) = H(x, t)$ is a path from y_0 to y_1.

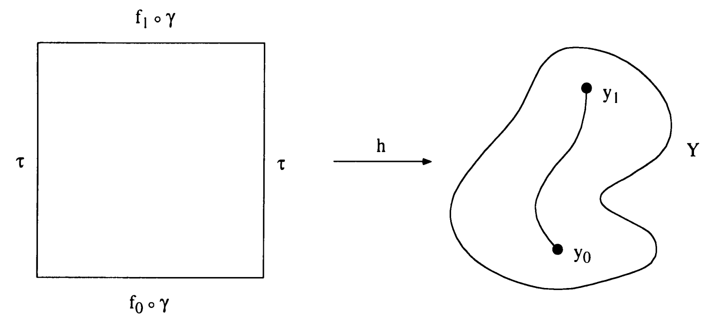
Proposition 12.11. *The diagram*



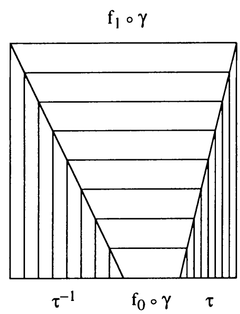
commutes, i.e., $\tau_\# \circ (f_0)_* = (f_1)_*$.

Proof. Let γ be a loop at x. We need to construct a homotopy from

the path $\tau^{-1}\cdot((f_0\circ\gamma)\cdot\tau)$ to the path $f_1\circ\gamma$. Consider the homotopy h from $[0,1]\times[0,1]$ to Y given by $h(t,s)=H(\gamma(t),s)$:



This provides a kind of homotopy between the path $\tau^{-1}\cdot((f_0\circ\gamma)\cdot\tau)$ around the two sides and the bottom of the square and the path $f_1\circ\gamma$ that goes across the top. □



Exercise 12.12. Turn this into a homotopy from $\tau^{-1}((f_0\circ\gamma)\cdot\tau)$ to $f_1\circ\gamma$ by constructing a continuous map from the square to itself that is the identity on the top side, is constant on the two sides, and maps the bottom side to the three sides, as indicated.

As a special case, we have the

Corollary 12.13. *If $f_0(x)=f_1(x)=y$, and H is a homotopy from f_0 to f_1 such that $H(x,s)=y$ for all s, then*

$$(f_0)_*=(f_1)_*\ :\ \pi_1(X,x)\ \rightarrow\ \pi_1(Y,y).$$

Two spaces X and Y are said to *have the same homotopy type* if there are continuous maps $f\colon X\rightarrow Y$ and $g\colon Y\rightarrow X$ such that $g\circ f$ is hom-

otopic to the identity map of X and $f \circ g$ is homotopic to the identity map of Y. The map f is called a *homotopy equivalence* if there is such a g.

Exercise 12.14. (a) Show that having the same homotopy type is an equivalence relation. (b) Show that a homotopy equivalence $f\colon X \to Y$ determines an isomorphism $f_*\colon \pi_1(X, x) \to \pi_1(Y, f(x))$ of fundamental groups. In particular, if $i\colon X \to Y$ embeds X as a deformation retract of Y, then $i_*\colon \pi_1(X, x) \to \pi_1(Y, i(x))$ is an isomorphism.

Exercise 12.15. Classify the following spaces according to homotopy type: (i) a point; (ii) a closed disk; (iii) a circle; (iv) $\mathbb{R}^2$; (v) $\mathbb{R}^n$; (vi) the complement of a point in the plane; (vii) two circles joined at a point; (viii) an annulus; and (ix) the complement of two points in the plane.

Exercise 12.16. Show that the mapping from S^2 to itself that takes (x, y, z) to $(-x, -y, z)$ is homotopic to the identity mapping, and the mapping that takes (x, y, z) to $(x, y, -z)$ is homotopic to the antipodal mapping.

Exercise 12.17. If $f\colon S^n \to S^n$ is a continuous mapping such that $f(P) \neq P$ for all P, show that f is homotopic to the antipodal mapping. If $f(P) \neq -P$ for all P, show that f is homotopic to the identity mapping.

Problem 12.18. If n is odd, show that the identity mapping on S^n is homotopic to the antipodal mapping.

It is a general fact that for even n, the antipodal map on S^n is not homotopic to the identity map. We will prove this in Chapter 23.

Problem 12.19. An orthogonal $(n+1) \times (n+1)$ matrix determines a mapping from S^n to itself. Show that two such mappings are homotopic if and only if they have the same determinant. If n is even, show that such a mapping is homotopic to the identity if the determinant is 1, and to the antipodal mapping if the determinant is -1.

Problem 12.20. Compute the fundamental group of the space $\mathrm{GL}_2^+(\mathbb{R})$ of (2×2)-matrices with positive determinant, which gets its topology as an open subspace of $\mathbb{R}^4$.

12c. Fundamental Group and Homology

For any topological space X, with base point x in X, there is a homomorphism from the fundamental group $\pi_1(X,x)$ to the homology group H_1X, that takes the class $[\gamma]$ of a loop γ at x to the homology class of γ, regarded as a closed path or 1-chain. It takes the constant path ε_x at x to 0. The fact that it is well defined amounts to the fact that homotopic paths define the same homology class, as we saw in Lemma 6.4. The same lemma showed that the homology class of a product $\sigma \cdot \tau$ of loops is the sum of the homology classes of the loops, which shows that the mapping is a homomorphism of groups.

Exercise 12.21. If $f: X \to Y$ is a continuous mapping, and $f(x) = y$, show that the following diagram commutes:

$$\begin{array}{ccc} \pi_1(X, x) & \xrightarrow{f_*} & \pi_1(Y, y) \\ \downarrow & & \downarrow \\ H_1X & \xrightarrow[f_*]{} & H_1Y. \end{array}$$

Since H_1X is an abelian group, this homomorphism must vanish on all commutators $a \cdot b \cdot a^{-1}b^{-1}$ in $\pi_1(X,x)$, so it must vanish on the *commutator subgroup* $[\pi_1(X,x), \pi_1(X,x)]$ that consists of all finite products of commutators. This is a normal subgroup of $\pi_1(X,x)$, and the quotient group is sometimes called the *abelianized fundamental group* of X, and denoted $\pi_1(X,x)_{\text{abel}}$. So we have a homomorphism

$$\pi_1(X,x)_{\text{abel}} = \pi_1(X,x)/[\pi_1(X,x), \pi_1(X,x)] \to H_1X.$$

Since the fundamental group $\pi_1(X,x)$ depends only on the path-connected component of X that contains x, we cannot expect the fundamental group to determine the homology group for disconnected spaces. But except for this, the fundamental group determines the homology group:

Proposition 12.22. *If X is a path-connected space, then the canonical homomorphism from $\pi_1(X,x)_{abel}$ to H_1X is an isomorphism.*

Proof. We must define a homomorphism from the abelian group Z_1X of 1-cycles to $\pi_1(X,x)_{\text{abel}}$, and show that the 1-boundaries B_1X map to zero. This will give a map back from H_1X to $\pi_1(X,x)_{\text{abel}}$. To define the map, let $\gamma = \Sigma_i n_i \gamma_i$ be a 1-cycle, with paths γ_i going from points $a(i)$ to $b(i)$ in X. For each point c that occurs as an endpoint of any

γ_i, choose a path τ_c from x to c. Let γ_i' be the loop at x defined by

$$\gamma_i' = \tau_{a(i)} \cdot \gamma_i \cdot \tau_{b(i)}^{-1},$$

where we are using the notation of Exercise 12.3(b). Define the map from Z_1X to $\pi_1(X,x)_{\text{abel}}$ by sending $\gamma = \Sigma_i n_i \gamma_i$ to the class of $\Pi_i[\gamma_i']^{n_i}$; note that the order in products is unimportant since the group $\pi_1(X,x)_{\text{abel}}$ is abelian.

We verify first that this is independent of the choice of paths τ_c. Suppose $\tilde{\tau}_c$ is another path from x to c, for each c. Let $\tilde{\gamma}_i' = \tilde{\tau}_{a(i)} \cdot \gamma_i \cdot \tilde{\tau}_{b(i)}^{-1}$. Let ϑ_c be the loop $\tilde{\tau}_c \cdot \tau_c^{-1}$. Then $[\tilde{\gamma}_i'] = [\vartheta_{a(i)}] \cdot [\gamma_i'] \cdot [\vartheta_{b(i)}^{-1}]$, so

$$\prod_i [\tilde{\gamma}_i']^{n_i} = \prod_i [\gamma_i']^{n_i} \cdot \left(\prod_i [\vartheta_{a(i)}]^{n_i} \cdot \prod_i [\vartheta_{b(i)}]^{-n_i} \right) = \prod_i [\gamma_i']^{n_i},$$

the last equation using the fact that γ is a 1-cycle, so each point c occurs as many times as a starting point $a(i)$ as ending point $b(i)$.

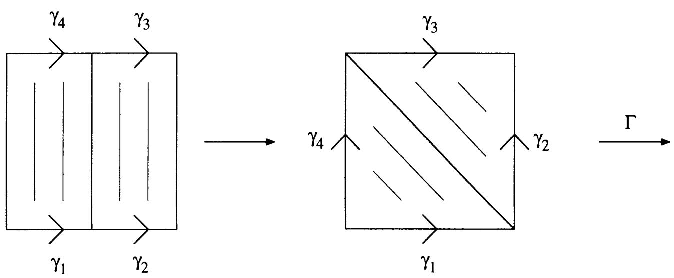
It follows from the definition that this map is a homomorphism from Z_1X to $\pi_1(X,x)_{\text{abel}}$. We next verify that this homomorphism maps boundary cycles to zero. It suffices to show that γ maps to zero when

$$\gamma = \partial\Gamma = \gamma_1 + \gamma_2 - \gamma_3 - \gamma_4,$$

where $\Gamma\colon [0,1] \times [0,1] \to X$ is a continuous mapping, and the boundary is as described in §6a. Let τ_1 and τ_2 be paths from x to the starting and ending points of γ_1, and let τ_3 and τ_4 be paths from x to the starting and ending points of γ_3. Then γ maps to the class of

$$\begin{aligned}&[\tau_1 \cdot \gamma_1 \cdot \tau_2^{-1}] \cdot [\tau_2 \cdot \gamma_2 \cdot \tau_4^{-1}] \cdot [\tau_4 \cdot \gamma_3^{-1} \cdot \tau_3^{-1}] \cdot [\tau_3 \cdot \gamma_4^{-1} \cdot \tau_1^{-1}] \\ &\qquad = [\tau_1 \cdot \gamma_1 \cdot \gamma_2 \cdot \gamma_3^{-1} \cdot \gamma_4^{-1} \cdot \tau_1^{-1}] = [(\tau_1 \cdot \gamma_1 \cdot \gamma_2) \cdot (\tau_1 \cdot \gamma_4 \cdot \gamma_3)^{-1}].\end{aligned}$$

To see that this last class is trivial in $\pi_1(X,x)$, it suffices to show that the paths $\tau_1 \cdot \gamma_1 \cdot \gamma_2$ and $\tau_1 \cdot \gamma_4 \cdot \gamma_3$ are homotopic with fixed endpoints. For this it is enough to show that $\gamma_1 \cdot \gamma_2$ and $\gamma_4 \cdot \gamma_3$ are homotopic with fixed endpoints. This is evident from the picture:



An explicit homotopy is given by the formula

$$H(t,s) = \begin{cases} \Gamma(2t(1-s), 2ts), & 0 \le t \le 1/2, \\ \Gamma((2t-1)s + 1 - s, (2t-1)(1-s) + s), & 1/2 \le t \le 1. \end{cases}$$

Finally, we must check that the two homomorphisms we have defined are inverse to each other. It is immediate from the definitions that the composite

$$\pi_1(X,x)_{\text{abel}} \to H_1X \to \pi_1(X,x)_{\text{abel}}$$

is the identity. For the other composite, for a 1-cycle $\gamma = \Sigma_i n_i \gamma_i$, and γ_i' defined as above, the element $\Pi_i[\gamma_i']^{n_i}$ in $\pi_1(X,x)_{\text{abel}}$ maps to

$$\sum_i n_i \gamma_1' = \sum_i n_i (\tau_{a(i)} + \gamma_i - \tau_{b(i)}) = \sum_i n_i \gamma_i = \gamma,$$

using the fact that the boundary of γ is zero. □

So far, all the fundamental groups we have calculated explicitly have been abelian, in which case the proposition says that $\pi_1(X,x) \cong H_1(X)$. We will soon calculate other examples with a nonabelian fundamental group. For now, however, we can appeal to Exercise 11.14 to find an example with a nonabelian fundamental group.

PART VII

COVERING SPACES AND FUNDAMENTAL GROUPS, II

In this part we will see that the two basic notions of covering spaces and fundamental groups are intimately related. In Chapter 13 we see how knowledge of the fundamental group controls the possible coverings a space may have: coverings correspond to subgroups of the fundamental group. There is a universal covering, from which all other coverings can be constructed. In Chapter 14 we use this the other way to prove the Van Kampen theorem, which relates the fundamental group of a union of two spaces to the fundamental groups of the two spaces and the fundamental group of their intersection. This can be regarded as the analogue for the fundamental group of the Mayer–Vietoris theorem for the homology group. The proof we give depends on a correspondence between G-coverings and homomorphisms from the fundamental group to G.

Caution to Beginners. The theorems relating coverings and the fundamental group are stated in their natural generality, with various technical conditions about the spaces involved. This has two advantages: you will have a reasonably complete story when you are finished, and the precise conditions sometimes help in shaping the proofs. However, this also has a large disadvantage: the proliferation of these technical conditions can get in the way of the main ideas. At least

for a first reading, it is probably a good idea to simply assume all spaces arising are reasonably behaved—for example, that they are open sets in the plane or a surface or manifold, or perhaps a finite graph. These cases, in fact, will suffice for the applications considered in this text.

CHAPTER 13

The Fundamental Group and Covering Spaces

13a. Fundamental Group and Coverings

We first have the basic:

Proposition 13.1. *If $p: Y \to X$ is a covering, and $p(y) = x$, then the induced homomorphism $p_*: \pi_1(Y, y) \to \pi_1(X, x)$ is an injection.*

Proof. This is a consequence of the lifting properties of Chapter 11. We must show that the kernel of p_* is $\{e\}$. If σ is a loop at y, and $p_*[\sigma] = e$, there is a homotopy H from $p \circ \sigma$ to the constant path ε_x at x. By homotopy lifting, H lifts to a homotopy $\tilde{H}$ from σ to some path. Since H maps the sides and top of the unit square to the point x, its lifting $\tilde{H}$ (by uniqueness of path-lifting) maps the sides and top of the square to the point y. So $\tilde{H}$ is a homotopy from σ to the constant path ε_y, and $[\sigma] = e$, as required. □

Exercise 13.2. (a) If σ is a loop at x, and $\tilde{\sigma}$ is the unique lifting of σ to a path starting at y, show that $\tilde{\sigma}$ ends at y if and only if its class is in the image of p_*, i.e., $[\sigma] \in p_*(\pi_1(Y, y))$. (b) If σ and σ' are two paths in X from x to x', and $\tilde{\sigma}$ and $\tilde{\sigma}'$ are the lifts to paths in Y starting at y, show that $\tilde{\sigma}$ and $\tilde{\sigma}'$ have the same endpoint if and only if $[\sigma' \cdot \sigma^{-1}]$ is in $p_*(\pi_1(Y, y))$. In the notation of §11b,

$$y*\sigma = y*\sigma' \iff [\sigma' \cdot \sigma^{-1}] \in p_*(\pi_1(Y, y)).$$

Exercise 13.3. If y' is another point in Y with $p(y') = x$, and y and y' can be connected by a path in Y, show that the image of $\pi_1(Y, y')$ in $\pi_1(X, x)$ is a subgroup conjugate to the image of $\pi_1(Y, y)$. In fact, if σ is a path from y' to y, and $\gamma = p \circ \sigma$ is its image, show that

$$p_*(\pi_1(Y, y')) = [\gamma] \cdot p_*(\pi_1(Y, y)) \cdot [\gamma]^{-1}.$$

Given a covering $p\colon Y \to X$, for any point y with $p(y) = x$, and any loop σ at x, we defined $y * \sigma$ to be the endpoint of the lift of σ that starts at y. This point $y * \sigma$ is also in $p^{-1}(x)$. We saw in §11b that if σ' is a loop homotopic to σ, then $y * \sigma' = y * \sigma$. For any homotopy class $[\sigma]$ in $\pi_1(X, x)$, we can therefore define $y * [\sigma]$ to be $y * \sigma$. This defines a right action of the fundamental group $\pi_1(X, x)$ on the fiber $p^{-1}(x)$:

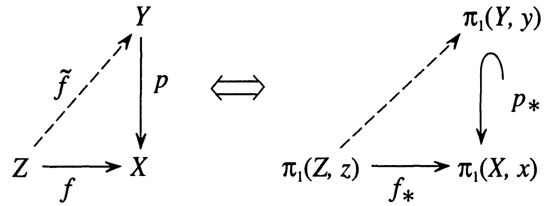
$$p^{-1}(x) \times \pi_1(X, x) \to p^{-1}(x), \qquad y \times [\sigma] \mapsto y * [\sigma],$$

taking $y \times [\sigma]$ to the endpoint of the lift of σ that starts at y.

Exercise 13.4. Show that this is a right group action; in particular, $y * ([\sigma] \cdot [\tau]) = (y * [\sigma]) * [\tau]$. If Y is path-connected, show that this action is transitive: for any y and y' in $p^{-1}(x)$ there is some $[\sigma]$ with $y * [\sigma] = y'$. Show that the subgroup that acts trivially on a point y is exactly $p_*(\pi_1(Y, y))$. For fixed y, show that this defines a one-to-one correspondence between set of right cosets $\pi_1(X, x)/p_*(\pi_1(Y, y))$ and the fiber $p^{-1}(x)$. In particular, the index of the subgroup $p_*(\pi_1(Y, y))$ in $\pi_1(X, x)$ is the number of sheets of the covering.

We saw in the last chapter that fundamental groups provide an obstruction to the existence of mappings: if there is no map between the groups, there cannot be a map between the spaces. The following proposition shows that, in the case of covering maps, the converse holds: if the fundamental groups say there may be a map, then there will be. Recall that a space X is *locally path-connected* if any neighborhood of any point contains a neighborhood that is path-connected.

Proposition 13.5. *Suppose $p\colon Y \to X$ is a covering, and $f\colon Z \to X$ is a continuous mapping with Z a connected and locally path-connected space. Let $x \in X$, $y \in Y$, $z \in Z$, be points with $p(y) = f(z) = x$. In order for there to be a continuous mapping $\tilde{f}$ $Z \to Y$ with $p \circ \tilde{f} = f$ and $\tilde{H}(z) = y$, it is necessary and sufficient that $f_*(\pi_1(Z, z))$ be contained in $p_*(\pi_1(Y, y))$:*



Such a lifting $\tilde{f}$, when it exists, is unique.

Proof. The necessity is clear from the functoriality of the fundamental group, and the uniqueness is a special case of Lemma 11.5. For the converse, to construct $\tilde{f}$, given a point w in Z, choose a path γ in Z from z to w, and let $\sigma = f \circ \gamma$, which is a path starting at x in X. Define $\tilde{f}(w)$ to be $y * \sigma$; that is, $\tilde{f}(w)$ is the endpoint of the path that lifts σ and starts at y. We must first show that this is independent of the choice of path. If γ' is another path from z to w, then $f \circ (\gamma' \cdot \gamma^{-1}) = \sigma' \cdot \sigma^{-1}$ is a loop at x. By the hypothesis, $[\sigma' \cdot \sigma^{-1}]$ is in the image of p_*. By Exercise 13.2(b) it follows that $\tilde{\sigma}'$ and $\tilde{\sigma}$ end at the same point.

We must verify that this mapping $\tilde{f}$ is continuous at the point w. Let N be any neighborhood of $f(w)$ that is evenly covered by p, let V be the open set in $p^{-1}(N)$ that maps homeomorphically onto N and that contains $\tilde{f}(w)$, and choose a path-connected neighborhood U of w so that $f(U) \subset N$. We need to show that $\tilde{f}$ maps U into V. For all points w' in U, we may find a path α from w to w' in U, and then we can use $\gamma \cdot \alpha$ as the path from z to w'. The lifting of $f \circ (\gamma \cdot \alpha) = (f \circ \gamma) \cdot (f \circ \alpha)$ is obtained by first lifting $f \circ \gamma$ to $\tilde{\sigma}$, then lifting $f \circ \alpha$. Since the latter lifting stays in V, this shows that $\tilde{f}(U) \subset V$. □

Corollary 13.6. *Let X be a connected and locally path-connected space. Let $p\colon Y \to X$ and $p'\colon Y' \to X$ be two covering maps, with Y and Y' connected and let $p(y) = x$ and $p'(y') = x$. In order for there to be an isomorphism between the coverings preserving the base points, it is necessary and sufficient that*

$$p_*(\pi_1(Y, y)) = p'_*(\pi_1(Y', y')).$$

Proof. The necessity is clear, and if these subgroups agree, the proposition gives maps $\varphi\colon Y \to Y'$ and $\psi\colon Y' \to Y$ preserving the base points and compatible with projections. (Note that X being locally path-connected implies that covering spaces Y and Y' are also locally path-connected.) Applying Lemma 11.5 to $\psi \circ \varphi$ and $\varphi \circ \psi$ shows that they are isomorphisms. □

Exercise 13.7. Show that two connected coverings of such an X are isomorphic, with an isomorphism that may not preserve base points, if and only if the images of their fundamental groups are conjugate subgroups of the fundamental group of X.

A path-connected space X is called *simply connected* if its fundamental group is the trivial group. Note that this is independent of choice of the base point. In the preceding proposition, if Z is simply connected, it follows that the liftings $\tilde{f}$ always exist.

Corollary 13.8. *A simply connected and locally path-connected space has only trivial coverings.*

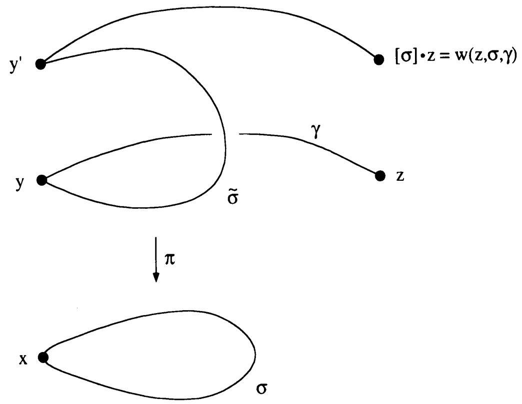
Proof. Suppose $p: Y \to X$ is a covering, with X simply connected. Fix x in X, and for each $y \in p^{-1}(x)$ apply the proposition to get a unique continuous map $s_y: X \to Y$ with $s_y(x) = y$ and $p \circ s_y$ the identity on X. This gives an isomorphism from the trivial covering $X \times p^{-1}(x)$ to Y, by $x \times y \mapsto s_y(x)$. □

Exercise 13.9. Verify that this map is a homeomorphism from the space $X \times p^{-1}(x)$ onto Y.

Problem 13.10. The hypotheses of locally path-connected are needed for the truth of Propositions 13.5 and the two corollaries. *Challenge.* Find an example of a path-connected space X with trivial fundamental group that has a nontrivial connected covering $p: Y \to X$.

13b. Automorphisms of Coverings

Our next goal is to relate the fundamental group of X to the automorphism group of a covering of X. Let $p: Y \to X$ be a covering, with base point y in Y chosen so that $p(y) = x$, and assume Y is path-connected. We want to make $\pi_1(X,x)$ act on the left on Y. Given an element $[\sigma]$ in $\pi_1(X,x)$ and a point z in Y, we therefore want to define a point $[\sigma] \cdot z$ in Y. Let $y' = y * [\sigma]$ be the endpoint of the lift of σ to a path that starts at y. Choose a path γ from y to z in Y. Since $p(y') = p(y) = x$, and $p \circ \gamma$ is a path starting at x, we have a point $y' * (p \circ \gamma)$ that is the endpoint of the lift of the path $p \circ \gamma$ that starts at y'. We want to define $[\sigma] \cdot z$ to be the point $y' * (p \circ \gamma)$. Denote this point $y' * (p \circ \gamma)$ temporarily by $w(z, \sigma, \gamma)$:



Equivalently, $w(z,\sigma,\gamma) = (y * \sigma) * (p \circ \gamma) = y * (\sigma \cdot (p \circ \gamma))$. Suppose, however, that we chose another path γ' from y to z. To have $w(z,\sigma,\gamma') = w(z,\sigma,\gamma)$, we want the lifts of $\sigma \cdot (p \circ \gamma')$ and $\sigma \cdot (p \circ \gamma)$ that start at y to end at the same point. By Exercise 13.2(b), this is the case precisely if the class $[(\sigma \cdot (p \circ \gamma')) \cdot (\sigma \cdot (p \circ \gamma))^{-1}]$ is in $p_*(\pi_1(Y,y))$. Note that $\gamma' \cdot \gamma^{-1}$ is a loop at y, and so

$$\begin{aligned}
[(\sigma \cdot (p \circ \gamma')) \cdot (\sigma \cdot (p \circ \gamma))^{-1}] &= [(\sigma \cdot (p \circ \gamma') \cdot (p \circ (\gamma^{-1})) \cdot \sigma^{-1})] \\
&= [\sigma] \cdot [p \circ (\gamma' \cdot \gamma^{-1})] \cdot [\sigma]^{-1} \\
&= [\sigma] \cdot p_*([\gamma' \cdot \gamma^{-1}]) \cdot [\sigma]^{-1},
\end{aligned}$$

which is an element of $[\sigma] \cdot p_*(\pi_1(Y,y)) \cdot [\sigma]^{-1}$. To know that this is in $p_*(\pi_1(Y,y))$, we need $p_*(\pi_1(Y,y))$ to be a *normal*[7] subgroup of $\pi_1(X,x)$. In this case we see that $w(z,\sigma,\gamma)$ is independent of choice of γ, and depends only on z and the homotopy class $[\sigma]$ of σ.

Assume then that $p_*(\pi_1(Y,y))$ is a normal subgroup of $\pi_1(X,x)$. The above construction determines a mapping

$$\pi_1(X,x) \times Y \to Y, \qquad [\sigma] \times z \mapsto [\sigma] \cdot z,$$

where $[\sigma] \cdot z = w(z,\sigma,\gamma)$. Note that $[\sigma] \cdot z$ is in the same fiber of p as z. Next we want to show that this is a left action of $\pi_1(X,x)$ on Y. The fact that $([\sigma] \cdot [\tau]) \cdot z = [\sigma] \cdot ([\tau] \cdot z)$ follows readily from the definition. For if γ is a path from y to z, then the lift of $\tau \cdot (p \circ \gamma)$, starting at y, is a path from y to $[\tau] \cdot z$. It follows that $[\sigma] \cdot ([\tau] \cdot z)$ is the endpoint of the lift of the path $\sigma \cdot (\tau \cdot (p \circ \gamma))$ that starts at y. The path $(\sigma \cdot \tau) \cdot (p \circ \gamma)$ is homotopic to $\sigma \cdot (\tau \cdot (p \circ \gamma))$, so its lift at y has the same endpoint, and this endpoint is $([\sigma] \cdot [\tau]) \cdot z$. The fact that $[\varepsilon_x] \cdot z = z$

[7] Recall that a subgroup H of a group G is a normal subgroup if $g \cdot H \cdot g^{-1} \subset H$ for all g in G.

follows from the fact that $[\varepsilon_x] \cdot z = (y * \varepsilon_x) * \gamma = y * \gamma$, and $y * \gamma = z$ by definition.

To prove that, for fixed $[\sigma]$, the map $z \mapsto [\sigma] \cdot z$ is continuous, we assume in addition that X is locally path-connected. To see the continuity near a point z, take, as in the preceding proposition, a path-connected neighborhood N of $p(z)$ that is evenly covered by p. Let V and V' be the components of $p^{-1}(N)$ that contain z and $z' = [\sigma] \cdot z$. We must show that $[\sigma] \cdot V$ is contained in V'. For v in V, let α be a path from z to v in V. If γ is a path from y and z, then $\gamma \cdot \alpha$ can be used as the path from y to v, from which it follows that $[\sigma] \cdot v$ is the endpoint of the lift of $p \circ \alpha$ that starts at z'. This lift is in V', which concludes the proof of continuity.

Summarizing, we have constructed a homomorphism from $\pi_1(X,x)$ to the group $\text{Aut}(Y/X)$ of covering transformations. We claim next that this is surjective. Let $\varphi\colon Y \to Y$ be a covering transformation, and suppose $\varphi(y) = y'$. It suffices to find an element $[\sigma]$ in $\pi_1(X,x)$ with $[\sigma] \cdot y = y'$, since, Y being connected, two covering transformations that agree at one point must be identical. Let γ be a path from y to y'. Let $\sigma = p \circ \gamma$. Then $[\sigma] \cdot y$ is the endpoint of the lift of the path σ that starts at y. Since this lift is γ, $[\sigma] \cdot y = y'$, as required.

Finally, we compute the kernel of this homomorphism from $\pi_1(X,x)$ to $\text{Aut}(Y/X)$. As in the preceding step, it suffices to see which $[\sigma]$ act trivially on y, and this happens when the lift of σ at y ends at y, i.e., when $[\sigma]$ is in $p_*(\pi_1(Y,y))$. Putting this all together, we have the:

Theorem 13.11. *Let $p\colon Y \to X$ be a covering, with Y connected and X locally path-connected, and let $p(y) = x$. If $p_*(\pi_1(Y,y))$ is a normal subgroup of $\pi_1(X,x)$, then there is a canonical isomorphism*

$$\pi_1(X,x)/p_*(\pi_1(Y,y)) \xrightarrow{\cong} \text{Aut}(Y/X).$$

The covering is a G-covering, with G being the quotient group $\pi_1(X,x)/p_*(\pi_1(Y,y))$.

The last statement follows from Proposition 11.38. A covering $p\colon Y \to X$ is called *regular* if $p_*(\pi_1(Y,y))$ is a normal subgroup of $\pi_1(X,x)$.

Exercise 13.12. Still assuming Y connected and X locally path-connected, but without assuming $H = p_*(\pi_1(Y,y))$ is a normal subgroup of $\pi_1(Y,y)$, let N be the normalizer of H in $\pi_1(X,x)$, i.e., N is the subgroup of elements g in $\pi_1(X,x)$ such that $g \cdot H \cdot g^{-1} \subset H$. Show that

N acts on the left on Y, and that this determines an isomorphism

$$N/H \cong \mathrm{Aut}(Y/X).$$

Exercise 13.13. Show that the following are equivalent (with Y connected and X locally path-connected): (i) the covering is regular; (ii) the action of $\mathrm{Aut}(Y/X)$ on $p^{-1}(x)$ is transitive; and (iii) for every loop σ at x, if one lifting of σ is closed, then all liftings are closed.

Exercise 13.14. Show that any G-covering $p: Y \to X$, with Y connected and locally path-connected, is a regular covering.

Corollary 13.15. *If $p: Y \to X$ is a covering, with Y simply connected and X locally path-connected, then $\pi_1(X,x) \cong Aut(Y/X)$.*

Corollary 13.16. *If a group G acts evenly on a simply connected and locally path-connected space Y, and $X = Y/G$ is the orbit space, then the fundamental group of X is isomorphic to G.*

In fact, we know by Proposition 11.37 that G is canonically isomorphic to the group of automorphisms of Y over X. Choosing a point y in Y over a point x in X determines an isomorphism of $\pi_1(X,x)$ with $\mathrm{Aut}(Y/X) = G$. □

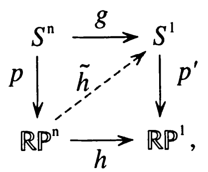
In general, the isomorphism of G with $\pi_1(X,x)$ depends on the choice of base point y, but only up to inner automorphism. In particular, in case the group is abelian, the isomorphism is independent of choices. For example, this corollary, applied to the mapping from $\mathbb{R}$ to S^1, implies again that $\pi_1(S^1,x) = \mathbb{Z}$.

Applied to the two-sheeted covering $p: S^n \to \mathbb{RP}^n$ from the sphere to the projective space, it follows that

$$\pi_1(\mathbb{RP}^n,x) = \mathbb{Z}/2\mathbb{Z} \qquad \text{for} \quad n \geq 2,$$

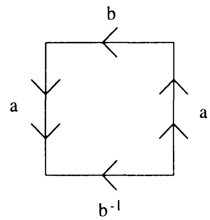
where x is any point in $\mathbb{RP}^n$.

This can be used to give a more conceptual explanation of something we saw in Chapter 4: if $n \geq 2$, there can be no continuous mapping from S^n to S^1 with $g(-P) = -g(P)$ for all P. Such a map would define a continuous mapping $h: \mathbb{RP}^n \to \mathbb{RP}^1$ on quotient spaces:



where $p'\colon S^1 \to \mathbb{RP}^1$ is the corresponding mapping for $n = 1$. Since $\pi_1(\mathbb{RP}^1, h(x)) \cong \mathbb{Z}$, the mapping $h_*\colon \pi_1(\mathbb{RP}^n, x) \to \pi_1(\mathbb{RP}^1, h(x))$ must be trivial. Choosing a point y in S^n with $p(y) = x$, by Proposition 13.5 there is a continuous mapping $\tilde{h}$ from $\mathbb{RP}^n$ to S^1 so that $p' \circ \tilde{h} = h$ and $\tilde{h}(x) = g(y)$. Now $\tilde{h} \circ p$ and g are two mappings from S^n to S^1 that map y to $g(y)$, and both, when followed by p', are the map $h \circ p$. By Lemma 11.5, $\tilde{h} \circ p = g$. But $\tilde{h} \circ p$ always takes P and $-P$ to the same point, while g never does. So such g cannot exist.

Exercise 13.17. (a) Compute the fundamental group of the Lens spaces of Exercise 11.30. (b) If the Klein bottle is constructed by identifying sides as shown:



show that the fundamental group has two generators a and b, with one relation $abab^{-1} = e$. In particular, the fundamental group is not abelian. (c) The torus is a two-sheeted covering of the Klein bottle, as in Exercise 11.27. Describe the image of the fundamental group of the torus in the fundamental group of the Klein bottle, and verify that it is a normal subgroup.

Exercise 13.18. If $p\colon Y \to X$ is the three-sheeted covering of Exercise 11.14, show that $p_*(\pi_1(Y, y))$ is not a normal subgroup of $\pi_1(X, x)$. In particular, $\pi_1(X, x)$ is not an abelian group.

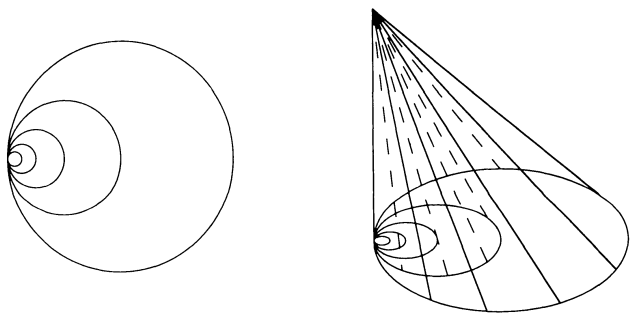
13c. The Universal Covering

In this section we assume that X is a connected and locally path-connected space. A covering $p\colon Y \to X$ is called a *universal covering* if Y is simply connected. It follows from Corollary 13.6 that such a covering, if it exists, is unique, and unique up to canonical isomorphism if base points are specified. As we have seen, the fundamental group of X will be isomorphic to the automorphism group of this

covering. The aim of this section is to show that, if one additional property is satisfied, a universal covering always exists.

Suppose we have a universal covering $p: Y \to X$. Any point in X has an evenly covered path-connected neighborhood N. Any loop σ in N lifts to a loop $\tilde{\sigma}$ in Y, and, since Y is simply connected, this loop is homotopic in Y to a constant path. It follows that the loop $\sigma = p \circ \tilde{\sigma}$ is homotopic to a constant path in X. A space X is called *semilocally simply connected* if every point has a neighborhood such that every loop in the neighborhood is homotopic in X to a constant path. So being semilocally simply connected is a necessary condition for the existence of a universal covering. Note that if X is *locally simply connected*, i.e., if every neighborhood of a point contains a neighborhood that is simply connected, then X is semilocally simply connected.

The spaces one generally meets, and those we have considered in this book, are all locally simply connected. Any open set in the plane or in $\mathbb{R}^n$, or any manifold, or any finite graph, is locally simply connected. In fact, one has to work a little to produce spaces that are not locally simply connected or semilocally simply connected.



Exercise 13.19. For a positive integer n, let C_n be the circle of radius $1/2^n$ centered at the point $(1/2^n, 0)$. Let $C \subset \mathbb{R}^2$ be the union of all the circles C_n; C is sometimes called a *clamshell*. (a) Show that C is connected and locally path-connected, but not semilocally simply connected. Let X be the cone over C, i.e., $X \subset \mathbb{R}^3$ is the union of all line segments from points in C to the point $(0, 0, 1)$. (b) Show that X is semilocally simply connected but not locally simply connected.

Suppose we have a universal covering $p: Y \to X$, with $p(y) = x$. For any point z in Y, there is a path γ from y to z, which is unique up to

homotopy since Y is simply connected. The image $\alpha = p \circ \gamma$ is a path from x to $p(z)$, unique up to homotopy (with fixed endpoints). Conversely, given a path α in X starting at x, it determines a point $z = y * \alpha$ in Y. This identifies Y, at least as a set, with the set of homotopy classes of paths in X that start at a given point x. The idea to the proof of the following theorem is to use this observation in reverse, by using these homotopy classes to construct the universal covering.

Theorem 13.20. *A connected and locally path-connected space X has a universal covering if and only if it is semilocally simply connected.*

Proof. To construct a universal covering, fix x in X and define $\tilde{X}$ to be the set of homotopy classes $[\gamma]$ of paths γ in X that start at x, the homotopies as usual being required to fix endpoints. Assigning to each such class its endpoint defines a function $u: \tilde{X} \to X$. Our task is to put a topology on $\tilde{X}$ so that this is a covering map, and show that $\tilde{X}$ is simply connected.

Call an open set N in X *good* if N is path-connected, and every loop in N at a point z in N is homotopic to the constant path ε_z in X. If γ is a path from x to a point z in N, let $N_{[\gamma]}$ be the subset of $\tilde{X}$ consisting of the homotopy classes $[\gamma \cdot \alpha]$, where α is any path in N that starts at z. Note that $N_{[\gamma]}$ depends only on the homotopy class $[\gamma]$ of γ. Here are some of the properties of these sets:

(1) if β is a path in a good N starting at the endpoint of γ, then $N_{[\gamma \cdot \beta]} = N_{[\gamma]}$;
(2) if a good N is contained in a good N', then $N_{[\gamma]} \subset N'_{[\gamma]}$; and
(3) if γ and γ' are two paths from x to a point z in a good N, then $N_{[\gamma]} = N_{[\gamma']}$ if γ and γ' are homotopic, and $N_{[\gamma]} \cap N_{[\gamma']} = \varnothing$ otherwise. □

Exercise 13.21. Verify these properties.

Now define a subset $\mathcal{O}$ of $\tilde{X}$ to be open if, for any $[\gamma]$ in $\mathcal{O}$, there is a good neighborhood N of the endpoint of γ with $N_{[\gamma]} \subset \mathcal{O}$. It follows from properties (1) and (2) that these open sets form a topology on $\tilde{X}$, and that each of these sets $N_{[\gamma]}$ is open, and the projection from $N_{[\gamma]}$ to N is continuous. In fact, this projection is a homeomorphism, the inverse being given by the map that takes w in N to $[\gamma \cdot \alpha]$, where α is any path in N from the endpoint of γ to w. This is independent of choice of α, since if α' is another, α and α' are homotopic in X, so $\gamma \cdot \alpha$ is homotopic to $\gamma \cdot \alpha'$. This projection is continuous since for smaller good N', N' maps to $N'_{[\gamma]}$.

The map $u: \tilde{X} \to X$ is evenly covered over any good N, since the

inverse image of N is a disjoint union of the open sets $N_{[\gamma]}$, where $[\gamma]$ varies over the homotopy classes of paths from x to any given point z of N.

For any path γ starting at x in X, and for s between 0 and 1, let γ_s be the path defined by $\gamma_s(t) = \gamma(st)$, $0 \le t \le 1$. The mapping $\tilde{\gamma}$: $[0, 1] \to \tilde{X}$ defined by $\tilde{\gamma}(s) = [\gamma_s]$ is the unique lift of γ to a path in $\tilde{X}$ that starts at the base point $\tilde{x} = [\varepsilon_x]$. In particular, it is a path in $\tilde{X}$ from $\tilde{x}$ to $[\gamma]$, showing that $\tilde{X}$ is connected. Any loop in $\tilde{X}$ at $\tilde{x}$ has the form $\tilde{\gamma}$ for a unique loop γ at x. For $\tilde{\gamma}$ to be a loop, the endpoint $\gamma_1 = \gamma$ must be homotopic to the constant path at x, which implies by the lifting of homotopies that $\tilde{\gamma}$ is homotopic to the constant path at $\tilde{x}$. This shows that $\tilde{X}$ is simply connected, and completes the proof of the theorem. □

Problem 13.22. (a) Suppose X is locally simply connected. Show that, if p: $Y \to X$ and q: $Z \to Y$ are covering maps, then $p \circ q$: $Z \to X$ is also a covering map. (b) Find a counterexample to (a) when X is a clamshell.

13d. Coverings and Subgroups of the Fundamental Group

The theorem of the preceding section will determine correspondence between subgroups of the fundamental group and coverings. For the following proposition, assume that X is connected, locally path-connected, and semilocally simply connected, so that X has a universal covering.

Proposition 13.23. (*a*) *For every subgroup H of $\pi_1(X,x)$ there is a connected covering p_H: $Y_H \to X$, with a base point $y_H \in p_H^{-1}(x)$ so that the image of $\pi_1(Y_H, y_H)$ in $\pi_1(X,x)$ is H. Any other such covering* (*with choice of base point*) *is canonically isomorphic to this one.*

(*b*) *If K is another subgroup of $\pi_1(X,x)$ containing H, there is a unique continuous mapping $p_{H,K}$: $Y_H \to Y_K$ that maps y_H to y_K and is compatible with the projections to X. This mapping is a covering mapping, and if H is a normal subgroup of K, it is a G-covering with $G = K/H$.*

Proof. Let u: $\tilde{X} \to X$ be a universal covering, with $u(\tilde{x}) = x$, and identify $\pi_1(X,x)$ with $\mathrm{Aut}(\tilde{X}/X)$. Any subgroup H of $\pi_1(X,x)$ acts evenly on $\tilde{X}$, and the quotient $\tilde{X} \to \tilde{X}/H = Y_H$ makes $\tilde{X}$ the universal cov-

ering of Y_H. The fundamental group of Y_H (with base point y_H the image of $\tilde{x}$) is canonically isomorphic to H. The projection from Y_H to $\tilde{X}/\pi_1(X,x) = X$ is a covering, and the image of its fundamental group in $\pi_1(X,x)$ is H. If H is contained in K, there is a canonical map on the orbit spaces $\tilde{X}/H \to \tilde{X}/K$. The remaining verifications are left to the reader, using the results of the preceding section. □

As seen in the proof, the covering $Y_H \to X$ corresponding to H can be identified with $\tilde{X}/H \to X$, where H acts on $\tilde{X}$ as a subgroup of $\mathrm{Aut}(\tilde{X}/X) = \pi_1(X,x)$. The regular coverings correspond to normal subgroups H. This means (see Exercise 13.14) that every connected G-covering $p\colon Y \to X$ has the form $\tilde{X}/H \to X$, with

$$\pi_1(X,x)/H \;\cong\; \mathrm{Aut}(Y/X) \;\cong\; G.$$

The correspondence of the proposition is similar to that seen in Galois theory, where subgroups correspond to field extensions, smaller subgroups corresponding to larger extensions. Here smaller subgroups of the fundamental group correspond to larger coverings of the space:

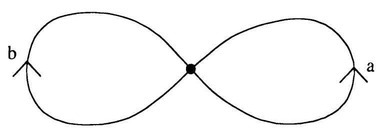
$$\begin{array}{cc} \tilde{X}, \tilde{x} & \{e\} \\ \downarrow & \cap \\ Y_H, y_H & H \\ \downarrow & \cap \\ Y_K, y_K & K \\ \downarrow & \cap \\ X, x & G \end{array}$$

If H is a normal subgroup of K, then the covering $Y_H \to Y_K$ is a K/H-covering.

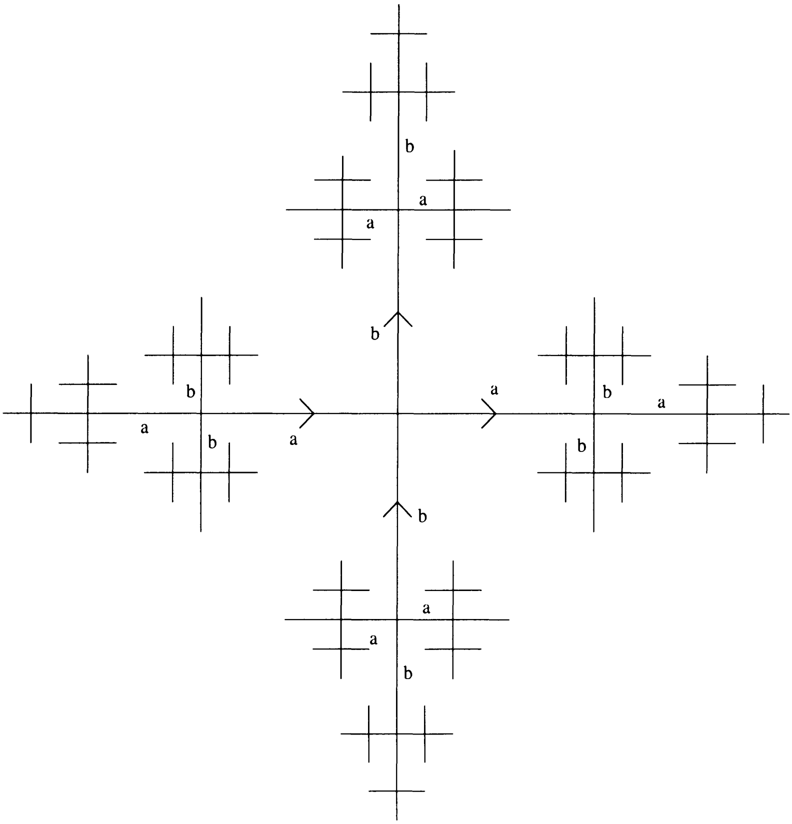
For example, if $X = S^1$ is a circle, the coverings correspond to subgroups of $\pi_1(S^1,(1,0)) = \mathbb{Z}$. The trivial group corresponds to the universal covering $\mathbb{R} \to S^1$, and the subgroups $\mathbb{Z}n \subset \mathbb{Z}$ correspond to the n-sheeted covering $p_n\colon S^1 \to S^1$ considered in §11a. Up to isomorphism, these are the only connected coverings of a circle.

Exercise 13.24. If $\pi_1(X,x)$ is abelian, show that every connected covering of X is regular.

One simple space whose universal covering we have not yet constructed is the figure 8 space



or the join of two circles at a point. This universal covering can be constructed as an infinite tree. Here part of it is pictured with segments of decreasing size, so that they don't overlap in the plane, but one should imagine that they are all of the same lengths.



Each segment maps to one of the loops of the figure 8 space, the horizontal segments to one and the vertical segments to the other.

Problem 13.25. (a) Show that this space is simply connected, and is the universal covering of the figure 8 space X. (b) If you know what a free group is, show that a free group with two generators acts evenly on this universal covering, with orbit space X. Deduce that the fundamental group of X is this free group. (c) Generalize to the space obtained by joining n circles at a point.

In Chapter 14 we will see a more general method of calculating

fundamental groups, which will include another proof of the result of the preceding problem.

The covering corresponding to the commutator subgroup of the fundamental group of X is a covering we may denote by $\tilde{X}_{\text{abel}}$:

$$\tilde{X}_{\text{abel}} = \tilde{X}/[\pi_1(X,x), \pi_1(X,x)] \to X.$$

This is a G-covering, with $G = \pi_1(X,x)/[\pi_1(X,x), \pi_1(X,x)]$, which is the first homology group H_1X by Proposition 12.22. It is sometimes called the *universal abelian covering* of X. In general a covering is called *abelian* if it is regular with abelian automorphism group.

Exercise 13.26. Show that any connected abelian covering of X has the form $\tilde{X}_{\text{abel}}/H$, for some subgroup H of H_1X.

We saw in Chapter 6 that for open sets in the plane, the homology class of a closed 1-chain is determined by winding numbers around points not in the set. The next exercise shows that this is true for other nice subsets, but the following problem shows it is not true in general.

Exercise 13.27. Suppose X is a subset of the plane that is contained in some open set U for which there is a retract $r\colon U \to X$. Show that a closed 1-chain γ on X is homologous to zero if and only if $W(\gamma, P) = 0$ for all P not in X.

Problem 13.28. *Challenge.* Let X be the clamshell of Exercise 13.19. Give an example of a closed path γ on X such that $W(\gamma, P) = 0$ for all P not in X, but such that γ is not a 1-boundary on X.

Exercise 13.29. The universal covering of the complement of the origin in the plane can be realized as the right half plane, via the polar coordinate mapping $(r, \vartheta) \mapsto (r\cos(\vartheta), r\sin(\vartheta))$, and as the entire complex plane $\mathbb{C}$ via the mapping $z \mapsto \exp(z)$. Find an isomorphism between these coverings.

CHAPTER 14

The Van Kampen Theorem

14a. G-Coverings from the Universal Covering

In this section X will denote a connected, locally path-connected, and semilocally simply connected space, so X has a universal covering, denoted $u: \tilde{X} \to X$. All spaces will have base points, and all maps will be assumed to take base points to base points. The base point of X is denoted x, and the base point of $\tilde{X}$ over x is denoted $\tilde{x}$.

We have seen in §13d that for every G-covering $p: Y \to X$, with Y connected, and with base point y, there is a surjective homomorphism of $\pi_1(X,x)$ onto G. If H is the kernel of this homomorphism, so $G \cong \pi_1(X,x)/H$, Y is the quotient of $\tilde{X}$ by the action of H, with y the image of $\tilde{x}$. We want to extend this correspondence to G-coverings that may not be connected. In this case there will only be a homomorphism from $\pi_1(X,x)$ to G, which need not be surjective. Here we will set up this correspondence between G-coverings and homomorphisms directly and rather briefly, omitting some verifications. Other ways to carry this out, with a more general context for these constructions, together with more details about the verifications, are described §16d and §16e.

Suppose $\rho: \pi_1(X,x) \to G$ is a homomorphism from the fundamental group of X to any group G. We will construct from ρ a G-covering $p_\rho: Y_\rho \to X$, together with a base point y_ρ in Y_ρ over x. Give G the discrete topology, so the Cartesian product $\tilde{X} \times G$ is a product of copies of $\tilde{X}$, one for each element in G. The group $\pi_1(X,x)$ acts on the left on $\tilde{X} \times G$ by the rule

$$[\sigma] \cdot (z \times g) = [\sigma] \cdot z \times g \cdot \rho([\sigma]^{-1}) = [\sigma] \cdot z \times g \cdot \rho([\sigma])^{-1},$$

for $[\sigma] \in \pi_1(X,x)$, $z \in \widetilde{X}$, $g \in G$. Here $[\sigma] \cdot z$ is the action of $\pi_1(X,x)$ on $\widetilde{X}$ that was described in §13b, and $g \cdot \rho([\sigma]^{-1})$ is the product in the group G. Define Y_ρ to be the quotient of $\widetilde{X} \times G$ by this action of $\pi_1(X,x)$:

$$Y_\rho = \widetilde{X} \times G/\pi_1(X,x),$$

and let y_ρ be the image of the point $\tilde{x} \times e$ in Y_ρ. Let $\langle z \times g\rangle$ denote the image in Y_ρ of the point $z \times g$ in $\widetilde{X} \times G$. Note that, by the above action of $\pi_1(X,x)$ on $\widetilde{X} \times G$, we have, for z in $\widetilde{X}$, g in G, and $[\sigma]$ in $\pi_1(X,x)$,

$$\langle [\sigma] \cdot z \times g\rangle = \langle z \times g \cdot \rho([\sigma])\rangle .$$

Define $p_\rho\colon Y_\rho \to X$ by taking $\langle z \times g\rangle$ to $u(z)$.

The group G acts on Y_ρ by the formula $h \cdot \langle z \times g\rangle = \langle z \times h \cdot g\rangle$, for h and g in G and z in X. (Note that using the right side of G for the left action of $\pi_1(X,x)$ frees up the left side of G for a left action of G!) We claim that this is an even action, making $p_\rho\colon Y_\rho \to X$ a G-covering. To prove this, let N be any open set in X over which the universal covering $u\colon \widetilde{X} \to X$ is trivial. By Lemma 11.18 there is an isomorphism of $u^{-1}(N)$ with the product covering $N \times \pi_1(X,x)$, on which $\pi_1(X,x)$ acts on the left on the second factor. This gives homeomorphisms

$$p_\rho^{-1}(N) \cong (N \times \pi_1(X,x)) \times G/\pi_1(X,x) \cong N \times G,$$

the latter homeomorphism by $\langle (u \times [\sigma]) \times g\rangle \mapsto u \times g \cdot \rho([\sigma])$. (The map back takes $u \times g$ to $\langle (u \times e) \times g\rangle$.) These homeomorphisms are compatible with the projections to N, and it follows that, over N, the action of G is even and the covering is a G-covering. Since X is covered by such open sets N, the same is true for the map p_ρ from Y_ρ to X.

Conversely, suppose $p\colon Y \to X$ is a G-covering, with a base point y over x. From this we construct a homomorphism ρ from $\pi_1(X,x)$ to G. For each $[\sigma]$ in $\pi_1(X,x)$ the element $\rho([\sigma])$ in G is determined by the formula

$$\rho([\sigma]) \cdot y = y * \sigma,$$

where $y * \sigma$ is the endpoint of the lift of the path σ that starts at y. We will need two facts about this operation:

(i) $(z * \sigma) * \tau = z * (\sigma \cdot \tau)$ for $z \in p^{-1}(x)$, σ a loop at x, and τ a path starting at x;

(ii) $g \cdot (z * \gamma) = (g \cdot z) * \gamma$ for $g \in G$, $z \in p^{-1}(x)$, and γ a path starting at x.

The first of these facts is immediate from the definition. The second follows from the fact that if $\tilde{\gamma}$ is lifting of γ starting at z, then the path $t \mapsto g \cdot \tilde{\gamma}(t)$, $0 \le t \le 1$, is a lifting of γ that starts at $g \cdot z$. The endpoint of this path, which is $(g \cdot z) * \gamma$ by definition, is $g \cdot \tilde{\gamma}(1)$, and since $\tilde{\gamma}(1) = z * \gamma$, (ii) follows.

We claim now that the ρ defined above is a homomorphism. This is a calculation, using (ii) and (i):

$$\begin{aligned}(\rho([\sigma]) \cdot \rho([\tau])) \cdot y &= \rho([\sigma]) \cdot (\rho([\tau]) \cdot y) = \rho([\sigma]) \cdot (y * \tau) \\ &= (\rho([\sigma]) \cdot y) * \tau = (y * \sigma) * \tau \\ &= y * (\sigma \cdot \tau) = \rho([\sigma] \cdot [\tau]) \cdot y.\end{aligned}$$

Proposition 14.1. *The above constructions determine a one-to-one correspondence between the set of homomorphisms from $\pi_1(X, x)$ to the group G and the set of G-coverings with base point, up to isomorphism*:

$$\mathrm{Hom}(\pi_1(X, x), G) \leftrightarrow \{G\text{-coverings}\}/\text{isomorphism}.$$

Proof. Given a G-covering $p\colon Y \to X$ with base points, from which we constructed a homomorphism ρ, we must now show that the given covering is isomorphic to the covering $p_\rho\colon Y_\rho \to X$ constructed from ρ. To map Y_ρ to Y, we need to map $\tilde{X} \times G$ to Y, and show that orbits by $\pi_1(X, x)$ have the same image. For this we identify the universal covering $\tilde{X}$ as the space of homotopy classes of paths in X starting at x. Define a map

$$\tilde{X} \times G \to Y, \qquad [\gamma] \times g \mapsto g \cdot (y * \gamma) = (g \cdot y) * \gamma.$$

This is easily checked to be continuous. We must check that an equivalent point $([\sigma] \cdot [\gamma]) \times (g \cdot \rho([\sigma])^{-1})$ maps to the same point. By (i) and (ii), this point maps to

$$\begin{aligned}(g \cdot \rho([\sigma])^{-1}) \cdot (y * (\sigma \cdot \gamma)) &= (g \cdot \rho([\sigma])^{-1}) \cdot ((y * \sigma) * \gamma) \\ &= ((g \cdot \rho([\sigma])^{-1}) \cdot (y * \sigma)) * \gamma \\ &= (g \cdot ((y * \sigma) * \sigma^{-1})) * [\gamma] \\ &= (g \cdot (y * (\sigma \cdot \sigma^{-1}))) * [\gamma] = (g \cdot y) * [\gamma],\end{aligned}$$

as required. Since the map takes the same values on equivalent points, it gives a mapping from the quotient Y_ρ to Y, which is a mapping of covering spaces of X. This is easily checked to be a mapping of G-coverings, from which it follows that it must be an isomorphism.

Conversely, starting with a homomorphism ρ, we constructed a G-

covering $Y_\rho \to X$, from which we constructed another homomorphism, say $\tilde{\rho}$. We must verify that $\tilde{\rho} = \rho$. Now for $[\sigma]$ in $\pi_1(X,x)$,

$$\begin{aligned}\tilde{\rho}([\sigma]) \cdot \langle \tilde{x} \times e \rangle &= \langle \tilde{x} \times e \rangle * \sigma = \langle \tilde{x} * \sigma \times e \rangle \\ &= \langle [\sigma] \cdot \tilde{x} \times e \rangle = \langle \tilde{x} \times e \cdot \rho([\sigma]) \rangle = \langle \tilde{x} \times \rho([\sigma]) \rangle \\ &= \rho([\sigma]) \cdot \langle \tilde{x} \times e \rangle .\end{aligned}$$

This shows that $\tilde{\rho}([\sigma]) = \rho([\sigma])$, which concludes the proof. □

Exercise 14.2. If $p\colon Y \to X$ is the G-covering corresponding to a homomorphism $\rho\colon \pi_1(X,x) \to G$, and X' is a subspace of X that also has a universal covering, with x in X', show that the restriction $p^{-1}(X') \to X'$ of this covering to X' is the G-covering corresponding to the composite homomorphism $\rho \circ i_*$, where $i_*\colon \pi_1(X',x) \to \pi_1(X,x)$ is induced by the inclusion i of X' in X.

Exercise 14.3. Show that, if base points are ignored, two G-coverings Y_ρ and $Y_{\rho'}$ are isomorphic G-coverings if and only if the homomorphisms ρ and ρ' are *conjugate*, i.e., there is some g in G such that

$$\rho'([\sigma]) = g \cdot \rho([\sigma]) \cdot g^{-1} \qquad \text{for all} \quad [\sigma] \in \pi_1(X,x).$$

14b. Patching Coverings Together

Suppose X is a union of two open sets U and V. A covering of X restricts to coverings of U and V, which are isomorphic over $U \cap V$. Conversely, suppose we have coverings $p_2\colon Y_1 \to U$ and $p_2\colon Y_2 \to V$, and we have an isomorphism of coverings

$$\vartheta\colon p_1^{-1}(U \cap V) \to p_2^{-1}(U \cap V)$$

of $U \cap V$. Then one may patch (or "glue," or "clutch") these together to get a covering $p\colon Y \to X$, together with isomorphisms of coverings

$$\varphi_1\colon Y_1 \xrightarrow{\cong} p^{-1}(U), \qquad \varphi_2\colon Y_2 \xrightarrow{\cong} p^{-1}(V)$$

of U and of V, so that, over $U \cap V$, $\vartheta = \varphi_2^{-1} \circ \varphi_1$.

One can construct Y as the quotient space of the disjoint union $Y_1 \sqcup Y_2$, by the equivalence relation that identifies a point y_1 in $p_1^{-1}(U \cap V)$ with the point $\vartheta(y_1)$ in $p_2^{-1}(U \cap V)$. (See Appendix A3.) Since ϑ is compatible with maps to X, one gets a mapping p from Y to X. Since the map from Y_1 to Y is a homeomorphism onto its image $p^{-1}U$, which is open in Y, and similarly Y_2 maps homeomorphically onto $p^{-1}V$, one sees that the restriction of p to the inverse image of

U is isomorphic to $Y_1 \to U$, and the restriction over V is isomorphic to $Y_2 \to V$. From this it follows in particular that p is a covering map.

If each of $Y_1 \to U$ and $Y_2 \to V$ is a G-covering, for a fixed group G, and ϑ is an isomorphism of G-coverings, then $Y \to X$ gets a unique structure of a G-covering in such a way that the maps from Y_1 and Y_2 commute with the action of G.

Occasionally the following generalization is useful. Suppose we have a collection X_α of open sets, $\alpha \in \mathcal{A}$, whose union is X, and a collection p_α: $Y_\alpha \to X_\alpha$ of covering maps. Suppose, for each α and β, we have an isomorphism

$$\vartheta_{\beta\alpha}\colon\ p_\alpha^{-1}(X_\alpha \cap X_\beta) \ \to\ p_\beta^{-1}(X_\alpha \cap X_\beta)$$

of coverings of $X_\alpha \cap X_\beta$. Assume these are compatible, i.e.,

(1) $\vartheta_{\alpha\alpha}$ is the identity on Y_α; and
(2) $\vartheta_{\gamma\alpha} = \vartheta_{\gamma\beta} \circ \vartheta_{\beta\alpha}$ on $p_\alpha^{-1}(X_\alpha \cap X_\beta \cap X_\gamma)$ for all α, β, $\gamma \in \mathcal{A}$.

Then one can patch these coverings together to obtain a covering p: $Y \to X$. One has isomorphisms φ_α: $Y_\alpha \to p^{-1}(X_\alpha)$ of coverings of X_α, such that $\vartheta_{\beta\alpha} = \varphi_\beta^{-1} \circ \varphi_\alpha$ on $p_\alpha^{-1}(X_\alpha \cap X_\beta)$. In addition, the space Y is the union of the open sets $\varphi_\alpha(Y_\alpha)$.

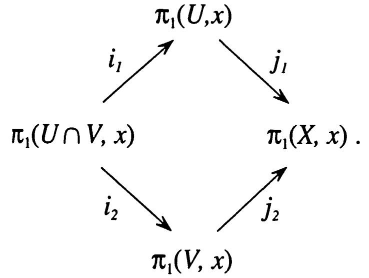
One constructs Y as the quotient space $\sqcup_{\alpha \in \mathcal{A}} Y_\alpha / R$ of the disjoint union of the Y_α by the equivalence relation determined by the $\vartheta_{\beta\alpha}$'s. The assertions about Y and the φ_α are general facts about patching spaces together, as proved in Appendix A3. The map p is determined by the equations $p \circ \varphi_\alpha = p_\alpha$ on Y_α. Since φ_α is a homeomorphism of Y_α onto $p^{-1}(X_\alpha)$, it follows that p is a covering map.

If each p_α: $Y_\alpha \to X_\alpha$ is a G-covering, with fixed G, and each $\theta_{\beta\alpha}$ is an isomorphism of G-coverings, then there is a unique action of G on Y so that each φ_α commutes with the action of G, i.e., $\varphi_\alpha(g \cdot y_\alpha) = g \cdot \varphi_\alpha(y_\alpha)$ for g in G and y_α in Y_α. This gives the patched covering p: $Y \to X$ the structure of a G-covering, so that each φ_α is an isomorphism of G-coverings.

14c. The Van Kampen Theorem

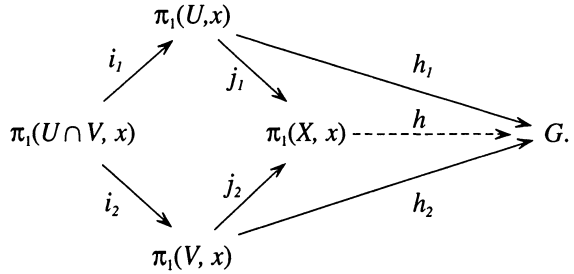
The Van Kampen theorem describes the fundamental group of a union of two spaces in terms of the fundamental group of each and of their intersection, under suitable hypotheses. Let X be a space that is a union of two open subspaces U and V. Assume that each of the spaces U, V and their intersection $U \cap V$ is path-connected, and let x be a

point in the intersection. Assume also that all these spaces X, U, V, and $U \cap V$ have universal covering spaces; this is the case, for example, if X is locally simply connected. We have a commutative diagram of homomorphisms of fundamental groups:



The maps are induced by the inclusions of subspaces, and commutativity means that $j_1 \circ i_1 = j_2 \circ i_2$.

We will describe how $\pi_1(X,x)$ is determined by the other groups (and the above maps between them). The description will not be direct, but will be by a *universal property*. Note that any homomorphism h from $\pi_1(X,x)$ to a group G determines a pair of homomorphisms $h_1 = h \circ j_1$ from $\pi_1(U,x)$ to G and $h_2 = h \circ j_2$ from $\pi_1(V,x)$ to G; the two homomorphisms $h_1 \circ i_1$ and $h_2 \circ i_2$ from $\pi_1(U \cap V, x)$ to G determined by these are the same. The Van Kampen theorem says that $\pi_1(X,x)$ is the "universal" group with this property.



Theorem 14.4 (Seifert–van Kampen). *For any homomorphisms*

$$h_1\colon \pi_1(U,x) \to G \quad \textit{and} \quad h_2\colon \pi_1(V,x) \to G,$$

such that $h_1 \circ i_1 = h_2 \circ i_2$, *there is a unique homomorphism*

$$h\colon \pi_1(X,x) \to G,$$

such that $h \circ j_1 = h_1$ *and* $h \circ j_2 = h_2$.

Exercise 14.5. Show that $\pi_1(X,x)$, together with the homomorphisms

j_1 and j_2, is determined up to canonical isomorphism by the universal property.

Exercise 14.6. Use the universal property to show that $\pi_1(X,x)$ is generated by the images of $\pi_1(U,x)$ and $\pi_1(V,x)$. Can you prove this assertion directly?

A version of the Van Kampen theorem was found first by Seifert, and the theorem is also known as the Seifert–Van Kampen theorem. The version given here, via universal properties, was given by Fox, see Crowell and Fox (1963). The usual proof of the Van Kampen theorem (without the hypotheses that the spaces all have universal coverings) is rather technical, and for it we refer to Crowell and Fox (1963) or Massey (1991). Here we will give a quick proof, due to Grothendieck (see Godbillon (1971)), using the correspondence between homomorphisms from fundamental groups to a group G and G-coverings. The assumptions assure that each of the spaces X, U, V, and $U \cap V$ has a universal covering space, and that homomorphisms from their fundamental groups to a group G correspond to G-coverings.

In particular, the homomorphisms h_1 and h_2 determine G-coverings $Y_1 \to U$ and $Y_2 \to V$, together with base points y_1 and y_2 over x. The fact that $h_1 \circ i_1$ is equal to $h_2 \circ i_2$ means that the restrictions of these coverings to $U \cap V$ are isomorphic G-coverings, and since $U \cap V$ is connected, there is a unique isomorphism between these G-coverings that maps the base point y_1 to the base point y_2. (The uniqueness is a special case of Exercise 11.24.) By the construction of the preceding section, these two coverings patch together, using this isomorphism over the intersection. This gives a G-covering $Y \to X$ that restricts to the two given G-coverings (and has the same base point). This G-covering corresponds to a homomorphism h from $\pi_1(X,x)$ to G, and the fact that the restricted coverings agree means precisely that $h \circ j_1 = h_1$ and $h \circ j_2 = h_2$. □

Corollary 14.7. *If U and V are simply connected, then X is simply connected.*

Note the important hypothesis in all these theorems, that all spaces, including the intersection $U \cap V$, are connected. It does not apply to the annulus, written as a union of two sets homeomorphic to disks!

Exercise 14.8. If V is simply connected, show that j_1: $\pi_1(U,x) \to \pi_1(X,x)$

is surjective, with kernel the smallest normal subgroup of $\pi_1(X, x)$ that contains the image of i_1: $\pi_1(U \cap V, x) \to \pi_1(U, x)$.

Corollary 14.9. *If $U \cap V$ is simply connected, then, for any G,*

$$\mathrm{Hom}(\pi_1(X, x), G) = \mathrm{Hom}(\pi_1(U, x), G) \times \mathrm{Hom}(\pi_1(V, x), G).$$

This means that $\pi_1(X, x)$ is the *free product* of $\pi_1(U, x)$ and $\pi_1(V, x)$.

Exercise 14.10. If $U \cap V$ is simply connected, show that the inclusion mappings j_1 and j_2 are one-to-one.

The following is a useful generalization of Van Kampen's theorem, which can be used to compute the fundamental group of an increasing union of spaces, each of whose fundamental groups in known. The proof is identical to that of the preceding theorem, using the general patching construction of the preceding section.

Suppose a space X is a union of a family of open subspaces X_α, $\alpha \in \mathscr{A}$, with the property that the intersection of any two of these subspaces is in the family. Assume that X and each X_α is path-connected and has a universal covering, and that the intersection of all the X_α contains a point x. When X_β is contained in X_α let $i_{\alpha\beta}$ be the map from $\pi_1(X_\beta, x)$ to $\pi_1(X_\alpha, x)$ determined by the inclusion, and let j_α be the map from $\pi_1(X_\alpha, x)$ to $\pi_1(X, x)$ determined by inclusion.

Theorem 14.11. *With these hypotheses, $\pi_1(X, x)$ is the* direct limit *of the groups $\pi_1(X_\alpha, x)$. That is, for any group G, and any collection of homomorphisms h_α from $\pi_1(X_\alpha, x)$ to G such that $h_\beta = h_\alpha \circ i_{\alpha\beta}$ whenever $X_\beta \subset X_\alpha$, there is a unique homomorphism h from $\pi_1(X, x)$ to G such that $h_\alpha = h \circ j_\alpha$ for all α.* □

The preceding theorem is recovered by taking the family to consist of U, V, and $U \cap V$.

Although this version of Van Kampen's theorem is stated with each subspace X_α open in X, it can often be applied to subspaces that are not open. For example, if each X_α is contained in an open set U_α, of which it is a deformation retract, with $U_\beta \subset U_\alpha$ whenever $X_\beta \subset X_\alpha$, and the hypotheses of Theorem 14.11 apply to these U_α, then $\pi_1(X, x)$ is the direct limit of the groups $\pi_1(X_\alpha, x)$. This follows from the fact that each $\pi_1(X_\alpha, x) \to \pi_1(U_\alpha, x)$ is an isomorphism. Without some such hypotheses, however, the theorem is false. For example, if A and B are copies of a cone over a clamshell (see Exercise 13.19), joined together at the one point where all the circles are tangent, then the

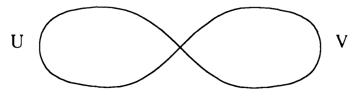
spaces A and B are simply connected, and $A \cap B$ is a point, but $A \cup B$ is not simply connected. (In fact, $A \cup B$ is an example of a space that is not simply connected but which has no nontrivial connected coverings.)

Exercise 14.12. Show if a space X is a union of a family of open subspaces X_α such that the intersection of any two sets in the family is also in the family, then H_1X is the direct limit of the groups H_1X_α.

14d. Applications: Graphs and Free Groups

One simple application of the Van Kampen theorem is a result we looked at earlier: the n-sphere S^n is simply connected if $n \geq 2$. To see this now, write the sphere as a union of two hemispheres each homeomorphic to n-dimensional disks, with the intersection homeomorphic to S^{n-1}. It follows from Corollary 14.7 that the fundamental group of S^n is trivial. (The assumption $n \geq 2$ is used to confirm that S^{n-1} is connected.)

Consider next a figure 8:



This is the union X of two circles U and V meeting at a point x. Let γ_1 and γ_2 be loops, one around each circle. The fundamental group of each circle is infinite cyclic, generated by the classes of these loops. It follows that to give a homomorphism from $\pi_1(X, x)$ to a group G is the same as specifying arbitrary elements g_1 and g_2 in G: there is a unique homomorphism from $\pi_1(X, x)$ to G mapping $[\gamma_1]$ to g_1 and $[\gamma_2]$ to g_2. This means that $\pi_1(X, x)$ is the *free group* on the generators $[\gamma_1]$ and $[\gamma_2]$.

Exercise 14.13. Let $a = [\gamma_1]$ and $b = [\gamma_2]$. Show that every element in $\pi_1(X, x)$ has a unique expression in the form

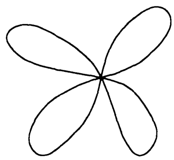
$$a^{m_0} \cdot b^{m_1} \cdot a^{m_2} \cdot \ . \ . \ . \ . \cdot b^{m_r},$$

where the m_i are integers, all nonzero except perhaps the first and last. The identity element is $e = a^0 b^0$.

The free group on two generators a and b can be constructed di-

rectly (and algebraically) as the set of all words of this form, with products defined by juxtaposition of words, canceling to get a legitimate word. It is straightforward (if a little awkward) to show by hand that this forms a group, and to see that it satisfies the above universal property. With the use of the Van Kampen theorem, we can avoid this, by constructing the free group as the fundamental group of this figure 8 space.

The free group F_n on n generators $a_1, \ldots, a_n$ is defined similarly: it is generated by these elements, and, for any group G and any elements $g_1, \ldots, g_n$ in G, there is a unique homomorphism from F_n to G taking a_i to g_i for $1 \leq i \leq n$. Again, it can be constructed purely algebraically using words in these letters, or as a fundamental group:



Exercise 14.14. Verify that the fundamental group of the space obtained by joining n circles at a point is the free group on n generators. Use this to show that the fundamental group of the complement of n points in the plane is free on n generators.

Exercise 14.15. Let X be a connected finite graph. (a) Show that, for any edge between two distinct vertices, X is homotopy equivalent to the graph obtained by removing the edge and identifying its two endpoints. (b) Show that X is homotopy equivalent to the graph obtained by joining n circles at a point, where, if the graph has v vertices and e edges, $n = e - v + 1$. (c) Show that n is the "connectivity" of the graph, i.e., the largest number of edges one can remove from the graph (leaving the vertices), so that what is left remains connected.

One can use this result to give a simple proof of a rather surprising fact about free groups.

Proposition 14.16. *If G is a free group on n generators, and H is a subgroup of G that has finite index d in G, then H is a free group, with $dn - d + 1$ generators.*

Proof. Take G to be the fundamental group of a connected graph X that has v vertices and e edges, with $n = e - v + 1$. For simplicity we

assume each edge of X connects two distinct vertices. The subgroup H corresponds to a connected covering $p: Y \to X$ with d sheets, with the fundamental group of Y isomorphic to H. It is not hard to verify that Y is a connected graph. In fact, the d points over each of the vertices of X can be taken as vertices of Y, and (since a covering is trivial over an interval), the d components of the inverse image of each edge are edges of Y. Since Y is a graph, its fundamental group is free, with

$$de - dv + 1 = d(e - v + 1) - d + 1 = dn - d + 1$$

generators. □

If U is the plane domain that is the complement of two points, then U has the figure 8 as a deformation retract. So U has the same fundamental group. In particular, this is not an abelian group. For example, the path $\gamma_1 \cdot \gamma_2 \cdot \gamma_1^{-1} \cdot \gamma_2^{-1}$ is not homotopic to a constant path (although all integrals of all closed 1-forms over this path are trivial).

Problem 14.17. Use the Van Kampen theorem to compute the fundamental groups of the complement of n points (or small disks) in: (1) a two-sphere; (2) a torus; and (3) a projective plane.

Problem 14.18. Describe the fundamental group of $\mathbb{R}^2 \setminus Z$, where Z is the set $\mathbb{Z}$ of all integers, or the set $\mathbb{Z}^2$ of lattice points, or any infinite discrete set.

Problem 14.19. Show that a free group on two generators contains a subgroup that is not finitely generated, in fact, a subgroup that is a free group on an infinite number of variables.

Problem 14.20. Use the Van Kampen theorem to compute the fundamental groups of: (1) the sphere with g handles; (2) the complement of n points in the sphere with g handles; and (3) the sphere with h crosscaps.

PART VIII

COHOMOLOGY AND HOMOLOGY, III

We have seen that coverings of a space can be described by giving coverings on open sets and patching together the coverings over the intersections of these open sets. In particular, one can start with trivial coverings over small open sets, and patch them together, and any covering arises this way. This process is formalized in Chapter 15, and the data that describe such coverings made into Čech cohomology classes. For G-coverings, when G is an abelian group, these classes form a group which we see is "dual" to the homology group, i.e., it is isomorphic to the group of homomorphisms from the first homology group into G.

This is applied to show that the first De Rham group H^1X of an open set in the plane is isomorphic to the dual $\mathrm{Hom}(H_1X, \mathbb{R})$ of the homology group, which gives the culmination of our experience that the homology group and the De Rham group are measuring the same thing about X. This allows us to translate results about homology into corresponding results about cohomology, and in particular to finish the proof of the Mayer–Vietoris theorem for cohomology. (Another proof of this fact will be given in the last chapter of this book.)

Chapter 16, which is optional, contains several miscellaneous variations, applications, and generalities on the same themes, with many of the details left as exercises. The patching construction is used to describe the orientation covering of a manifold. The construction of a covering of a plane domain from a closed 1-form, which follows from the general results of Chapter 15, is carried out here directly using the language of "germs" of functions; in particular, the cov-

erings are seen as graphs of multivalued functions, similar to the polar coordinate covering of Part I. We also describe briefly another cohomology theory.

The last few sections of Chapter 16 generalize the constructions relating coverings with actions of groups and group homomorphisms that were used in Chapters 14 and 15. The added generality, while not needed in this book, may help to make the constructions more understandable by putting them in their natural context. In addition, the exercises (with their hints) carry out proofs of general facts about these constructions, special cases of which were used in Part VII.

CHAPTER 15

Cohomology

15a. Patching Coverings and Čech Cohomology

Since a G-covering is locally trivial, it can be constructed by patching together trivial coverings. In this section we specify exactly what data are needed to carry this out. First, we need to know what this pasting data looks like.

Lemma 15.1. *If $Y = X \times G \to X$ is the trivial G-covering, and $h: X \to G$ is any locally constant function, then the mapping*

$$x \times g \mapsto x \times g \cdot h(x)$$

from Y to Y is an isomorphism of G-coverings, and every isomorphism of G-coverings has this form for a unique locally constant function h.

Proof. It is evident that such a map is continuous and compatible with left multiplication by G, so it is an isomorphism of G-coverings. Conversely, if $\varphi: Y \to Y$ is compatible with left multiplication by G, φ must take each point $x \times g$ to $x \times g \cdot h(x)$ for some $h(x)$ in G. For φ to be continuous, the map $x \mapsto h(x)$ must be locally constant. □

If $p: Y \to X$ is any G-covering, one can find a collection $\mathfrak{U} = \{U_\alpha : \alpha \in \mathcal{A}\}$ of open sets whose union is X such that the restriction of the covering to each U_α is a trivial G-covering. Choose isomorphisms of G-coverings

$$\varphi_\alpha : U_\alpha \times G \xrightarrow{\cong} p^{-1}(U_\alpha).$$

On the overlaps $U_\alpha \cap U_\beta$ that are not empty we have "transition" isomorphisms

$$U_\alpha \cap U_\beta \times G \rightarrow p^{-1}(U_\alpha \cap U_\beta) \rightarrow U_\alpha \cap U_\beta \times G$$

given by the restriction of φ_α followed by the restriction of ${\varphi_\beta}^{-1}$. These transition isomorphisms have the form $x \times g \mapsto x \times g \cdot g_{\alpha\beta}(x)$ for some (unique) locally constant functions $g_{\alpha\beta}\colon U_\alpha \cap U_\beta \rightarrow G$, by the preceding lemma. The collection of these $g_{\alpha\beta}$ satisfy three properties:

(i)	$g_{\alpha\alpha} = e$	(the identity in G) for all α;
(ii)	$g_{\beta\alpha} = (g_{\alpha\beta})^{-1}$	for all α, β; and
(iii)	$g_{\alpha\beta} = g_{\alpha\beta} \cdot g_{\beta\gamma}$	on $U_\alpha \cap U_\beta \cap U_\gamma$ for all α, β, γ.

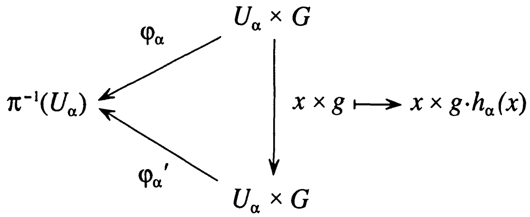
The first two of these are obvious, and the last follows from the equation ${\varphi_\gamma}^{-1} \circ \varphi_\alpha = {\varphi_\gamma}^{-1} \circ \varphi_\beta \circ {\varphi_\beta}^{-1} \circ \varphi_\alpha$. A collection $\{g_{\alpha\beta}\}$ of locally constant functions[8] satisfying (i)–(iii) is called a *Čech cocycle on* $\mathcal{U}$ *with coefficients in* G.

Conversely, given $\mathcal{U}$ and a Čech cocycle $\{g_{\alpha\beta}\}$ on $\mathcal{U}$ with coefficients in G, one can use it as "gluing data" to construct a G-covering, together with a trivialization over each U_α, so that the resulting cocycle is the given one. To do this, take the disjoint union of all the products $U_\alpha \times G$ (where G has the discrete topology, i.e., its points are open), and define an equivalence relation by defining $x \times g$ in $U_\alpha \times G$ to be equivalent to $x \times g \cdot g_{\alpha\beta}(x)$ in $U_\beta \times G$ for x in $U_\alpha \cap U_\beta$. Properties (i)–(iii) guarantee precisely that this is an equivalence relation. Define Y to be the set of equivalence classes, with the quotient topology. Since the equivalence relation is compatible with left multiplication by G and with the projections to X, the space Y gets an action of G and a projection p from Y to X so that the resulting maps $U_\alpha \times G \rightarrow Y$ are G-maps and compatible with the maps to X. In other words, this is the patching described in §14b, using the transition functions $\vartheta_{\beta\alpha}$, where $\vartheta_{\beta\alpha}(x \times g) = x \times g \cdot g_{\alpha\beta}(x)$.

The cocycle is not uniquely determined by the G-covering, even with a fixed choice of $\mathcal{U}$, since it depends on the choice of trivializations φ_α. But if $\{\varphi_\alpha'\}$ is another choice of trivializations, using the lemma again, there is for each α a unique locally constant function

[8] Note that if the collection U is chosen so that every $U_\alpha \cap U_\beta$ is connected, then these $g_{\alpha\beta}$ are constant, i.e., they are just elements of G. This case will suffice for our applications, and the reader is invited to make this simplifying assumption from the start.

$h_\alpha: U_\alpha \to G$ so that the diagram



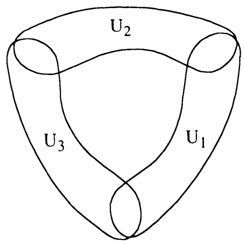
commutes. From this it follows that the cocycle $\{g_{\alpha\beta}'\}$ for the trivializations φ_α' is related to that $\{g_{\alpha\beta}\}$ for the φ_α by the equations

$$g_{\alpha\beta}' = (h_\alpha)^{-1} \cdot g_{\alpha\beta} \cdot h_\beta .$$

Two Čech cocycles $\{g_{\alpha\beta}\}$ and $\{g_{\alpha\beta}'\}$ are said to be *cohomologous* if there are locally constant functions $h_\alpha: U_\alpha \to G$ such that $g_{\alpha\beta}' = (h_\alpha)^{-1} \cdot g_{\alpha\beta} \cdot h_\beta$ on $U_\alpha \cap U_\beta$ for all α, β such that $U_\alpha \cap U_\beta$ is nonempty. This is easily checked to be an equivalence relation, and the equivalence classes are called *Čech cohomology classes* on $\mathcal{U}$ *with coefficients in* G. The set of these is denoted $H^1(\mathcal{U}; G)$.

Exercise 15.2. Verify that the G-coverings constructed from cohomologous cocycles are isomorphic.

Exercise 15.3. Let $\mathcal{U}$ be a covering of an annulus by three sets homeomorphic to disks, as shown:



For g in G let $c(g)$ be the cocycle determined by setting $g_{12} = g_{23} = e$ and $g_{31} = g$. (a) Show that every Čech cocycle on $\mathcal{U}$ with coefficients in G is cohomologous to such a cocycle, and show that $c(g)$ is cohomologous to $c(g')$ if and only if g and g' are conjugate in G. This gives a bijection between the set of conjugacy classes in G and $H^1(\mathcal{U}; G)$. (b) If G is cyclic of prime order, show that the coverings corresponding to any two distinct elements that are not the identity in G are isomorphic as coverings, but not as G-coverings.

What we have done in this section amounts to setting up a one-to-one correspondence between $H^1(\mathcal{U};G)$ and

$$\{G\text{-coverings of } X \text{ that are trivial over each } U_\alpha\}/\cong.$$

If the U_α are chosen to be simple open sets like disks or rectangles, or any simply connected and locally path-connected open sets, then every covering is trivial over them (by Corollary 13.8), so we will have all G-coverings of X classified by Čech cohomology classes. Thus we have:

Proposition 15.4. *If each U_α is simply connected and locally path-connected, then the set of G-coverings of X, up to isomorphism, is in one-to-one correspondence with the Čech cohomology set $H^1(\mathcal{U};G)$.*

Now suppose X is a connected space that has a universal covering. From Exercise 14.3 we know that the set of G-coverings up to isomorphism is in one-to-one correspondence with the set of homomorphisms from $\pi_1(X,x)$ to G, up to conjugacy. Putting this together, and continuing to assume that the open sets in $\mathcal{U}$ are simply connected, we have

Corollary 15.5. *With these assumptions, there is a bijection*

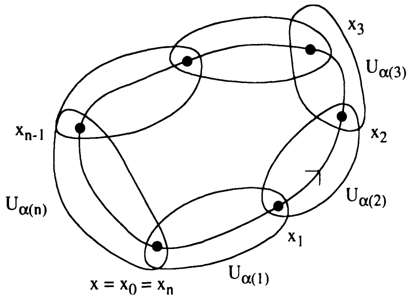
$$H^1(\mathcal{U};G) \leftrightarrow \operatorname{Hom}(\pi_1(X,x),G)/\textit{conjugacy}.$$

15b. Čech Cohomology and Homology

If we want to have a well-defined homomorphism from $\pi_1(X,x)$ to G, not just up to conjugacy, we know that we must consider coverings $Y \to X$ together with a base point y in Y over x. To do this with the patching, for those open sets U_α that contain x, fix the local trivializations $\varphi_\alpha\colon U_\alpha \times G \to p^{-1}(U_\alpha)$ so that $\varphi_\alpha(x \times e) = y$. The corresponding cocycle then has $g_{\alpha\beta}(x) = e$ whenever x is in $U_\alpha \cap U_\beta$; another cocycle defines an isomorphic covering preserving base points when it is of the form $(h_\alpha)^{-1} \cdot g_{\alpha\beta} \cdot h_\beta$, with $h_\alpha\colon U_\alpha \to G$ locally constant with the restriction that $h_\alpha(x) = e$ whenever x is in U_α. Such equivalence classes form a set we can denote by $H^1(\mathcal{U},x;G)$. This sets up bijections

$$\begin{aligned} H^1(\mathcal{U},x;G) &\leftrightarrow \{G\text{-coverings with base point}\}/\cong \\ &\leftrightarrow \operatorname{Hom}(\pi_1(X,x),G). \end{aligned}$$

We will need a prescription to make the homomorphism ρ from $\pi_1(X, x)$ to G explicit, when the covering is given as above by a cocycle. As one should expect, the answer is the product of values of the transition functions as one moves around a path, just as the winding number is the sum of changes in angle. If γ is a loop at x, subdivide the unit interval, $0 = t_0 < t_1 < \ldots < t_n = 1$ such that γ maps each subinterval $[t_{i-1}, t_i]$ into one of the open sets $U_{\alpha(i)}$. Let $x_i = \gamma(t_i)$.



Lemma 15.6. *The homomorphism $\rho: \pi_1(X, x) \to G$ corresponding to the G-covering $p: Y \to X$ takes the homotopy class of γ to*

$$\rho([\gamma]) = h_1 \cdot h_2 \cdot \ldots \cdot h_{n-1} \cdot h_n,$$

where $h_i = g_{\alpha(i)\alpha(i+1)}(x_i)$ for $1 \leq i \leq n-1$, and $h_n = g_{\alpha(n)\alpha(1)}(x_n)$.

Proof. Let $\tilde{\gamma}$ be the lift of γ starting at y. We must compute $\tilde{\gamma}(t_i)$ for each i. At the start, $\tilde{\gamma}(0) = y = \varphi_{\alpha(1)}(x \times e)$. For $0 \leq t \leq t_1$, by continuity, $\tilde{\gamma}(t) = \varphi_{\alpha(1)}(\gamma(t) \times e)$. So

$$\tilde{\gamma}(t_1) = \varphi_{\alpha(1)}(x_1 \times e) = \varphi_{\alpha(2)}(x_1 \times e \cdot g_{\alpha(1)\alpha(2)}(x_1)) = \varphi_{\alpha(2)}(x_1 \times h_1).$$

Going along each piece of the path in the same way, we find

$$\tilde{\gamma}(t_i) = \varphi_{\alpha(i)}(x_i \times h_1 \cdot \ldots \cdot h_{i-1}) = \varphi_{\alpha(i+1)}(x_i \times h_1 \cdot \ldots \cdot h_i).$$

At the end, this gives

$$\begin{aligned}\tilde{\gamma}(1) &= \varphi_{\alpha(1)}(x \times h_1 \cdot \ldots \cdot h_n) = (h_1 \cdot \ldots \cdot h_n) \cdot \varphi_{\alpha(1)}(x \times e) \\ &= (h_1 \cdot \ldots \cdot h_n) \cdot y.\end{aligned}$$

Since $\tilde{\gamma}(1) = \rho([\gamma]) \cdot y$ by definition, the lemma follows. □

Now we specialize to the case where G is an abelian group. In this case any homomorphism from $\pi_1(X, x)$ to G must send any commutator $aba^{-1}b^{-1}$ in $\pi_1(X, x)$ to the identity of G, so it must define a

homomorphism on the abelian quotient group $\pi_1(X,x)_{\text{abel}}$. By Proposition 12.22, we have an isomorphism $\pi_1(X,x)_{\text{abel}} \cong H_1X$. And conjugate homomorphisms to an abelian group must be equal. Putting all this together, and assuming X and $\mathcal{U}$ are as in Proposition 15.4, we have:

Corollary 15.7 (Hurewicz). *If G is abelian, there are canonical bijections*

$$\begin{aligned} H^1(\mathcal{U};G) &\leftrightarrow \operatorname{Hom}(\pi_1(X,x),G) \leftrightarrow \{G\text{-coverings of } X\}/\cong \\ &\leftrightarrow \operatorname{Hom}(\pi_1(X,x)/[\pi_1(X,x),\pi_1(X,x)],G) \\ &\leftrightarrow \operatorname{Hom}(H_1X,G). \end{aligned}$$

For a space X that may not be connected, we have:

Corollary 15.8. *If X is a locally simply connected space, and $\mathcal{U}$ is a collection of simply connected open sets whose union is X, then there are canonical bijections*

$$H^1(\mathcal{U};G) \leftrightarrow \{G\text{-coverings of } X\}/\cong \leftrightarrow \operatorname{Hom}(H_1X,G).$$

Proof. To give a G-covering of X is the same as giving a G-covering of each connected component of X, and to give a homomorphism from H_1X to G is the same as giving a homomorphism from the first homology group of each connected component of X to G. So the result follows from Corollary 15.7. □

If G is abelian, the Čech cocycles on $\mathcal{U}$ with coefficients in G form an abelian group, by multiplication: $\{g_{\alpha\beta}\}\cdot\{g_{\alpha\beta}'\} = \{g_{\alpha\beta}\cdot g_{\alpha\beta}'\}$. Call a cocycle $\{g_{\alpha\beta}\}$ a *coboundary* if there are $h_\alpha: U_\alpha \to G$ for each α such that $g_{\alpha\beta} = h_\alpha^{-1}\cdot h_\beta$.

Exercise 15.9. Assume that G is abelian. Show that the coboundaries form a subgroup of the cocycles, and that $H^1(\mathcal{U};G)$ is the quotient group of cocycles modulo coboundaries. (Caution: If G is not abelian, $H^1(\mathcal{U};G)$ has no natural group structure, see Exercise 15.3.) If G is abelian, give $\operatorname{Hom}(H,G)$ the structure of an abelian group, for any group H. Show that the bijections of Corollary 15.7 are isomorphisms of abelian groups.

Exercise 15.10. Define the 0th Čech group $H^0(\mathcal{U};G)$ for an open covering $\mathcal{U}$ of a space X with coefficients in a group by defining a class (or cocycle—there is no equivalence relation) to be a collection of locally constant functions $g_\alpha: U_\alpha \to G$ such that $g_\alpha = g_\beta$ on $U_\alpha \cap U_\beta$.

Show that $H^0(\mathcal{U}; G)$ is the direct product of copies of G, one for each connected component of X. In particular, for X open in the plane, $H^0(\mathcal{U}; \mathbb{R})$ is isomorphic to the space $H^0(X)$ of locally constant functions.

15c. De Rham Cohomology and Homology

We want to compare the De Rham group H^1X with the first homology group H_1X, when X is an open set in the plane. If ω is a closed 1-form, and γ a closed 1-chain, we defined the integral $\int_\gamma \omega$ in Chapter 9. For fixed ω, this map $\gamma \mapsto \int_\gamma \omega$ is a homomorphism from the group Z_1X of closed 1-chains to $\mathbb{R}$. Proposition 9.11 says that this integral is the same for homologous 1-chains. In other words, the map vanishes on 1-boundaries B_1X, and therefore defines a homomorphism from H_1X to $\mathbb{R}$. Let us denote this homomorphism, at least temporarily, by φ_ω, so that $\varphi_\omega([\gamma]) = \int_\gamma \omega$.

The set $\mathrm{Hom}(H_1X, \mathbb{R})$ of homomorphisms from H_1X to $\mathbb{R}$ has a natural structure of vector space: the sum $\varphi + \psi$ of two homomorphisms is defined by $(\varphi + \psi)([\gamma]) = \varphi([\gamma]) + \psi([\gamma])$, and multiplication of φ by a scalar r by $(r \cdot \varphi)([\gamma]) = r \cdot \varphi([\gamma])$. (Note that this works for any group in place of H_1X.) The above map that assigns φ_ω to the closed 1-form ω is a linear mapping of vector spaces

$$\{\text{closed 1-forms on } X\} \to \mathrm{Hom}(H_1X, \mathbb{R}).$$

This follows from the equation $\int_\gamma (r_1\omega_1 + r_2\omega_2) = \int_\gamma r_1\omega_1 + \int_\gamma r_2\omega_2$. If the 1-form ω is exact, the homomorphism φ_ω is zero. In fact, if $\omega = df$, and γ is any 1-chain, with boundary $\partial\gamma = \Sigma m_i P_i$, then

$$\int_\gamma df = \sum m_i f(P_i).$$

It follows that the above mapping vanishes on the subspace of exact 1-forms, so it defines a linear map on the quotient space H^1X. That is, we have a natural linear map of vector spaces

$$H^1X \to \mathrm{Hom}(H_1X, \mathbb{R}), \qquad [\omega] \mapsto \left([\gamma] \mapsto \int_\gamma \omega\right).$$

If $\alpha \in H^1X$ and $a \in H_1X$, we may write simply $\int_a \alpha$ in place of $\int_\gamma \omega$, where ω is a representative of α and γ a representative of a. The goal of this section is to show that this mapping is always an isomorphism.

Theorem 15.11. *If X is an open set in the plane, then the canonical homomorphism*

$$H^1X \to \operatorname{Hom}(H_1X, \mathbb{R})$$

is an isomorphism.

The fact that this map is one-to-one is not hard to see, for if a closed 1-form ω has all integrals $\int_\gamma \omega$ vanishing for all closed paths γ on X, we know from Proposition 1.8 that ω is exact. The fact that every homomorphism comes from a 1-form, however, will take some work. As a warm-up you may consider the special case:

Exercise 15.12. Show that the homomorphism $H^1X \to \operatorname{Hom}(H_1X, \mathbb{R})$ is an isomorphism when X is multiply connected.

Exercise 15.13. Show that, if U is an open subset of X, the diagram

$$\begin{array}{ccc} H^1X & \longrightarrow & H^1U \\ \downarrow & & \downarrow \\ \operatorname{Hom}(H_1X, \mathbb{R}) & \longrightarrow & \operatorname{Hom}(H_1U, \mathbb{R}) \end{array}$$

commutes, where the map on the bottom is determined by the map from H_1U to H_1X.

In order to prove the theorem, we specialize the results of §15b to the case where $G = \mathbb{R}$ is the additive group of real numbers, and X is an open set in the plane. Let $\mathcal{U} = \{U_\alpha : \alpha \in \mathcal{A}\}$ be a collection of open rectangles whose union is X. We may find such a collection so that any point in X is contained in only finitely many U_α (see Lemma A.20 and Lemma 24.10). By Corollary 15.8 we have a bijection between $\operatorname{Hom}(H_1X, \mathbb{R})$ and $H^1(\mathcal{U}; \mathbb{R})$. To prove the theorem we will construct a map from $H^1(\mathcal{U}; \mathbb{R})$ to the De Rham group H^1X, and then show that all these maps are compatible and isomorphisms.

An element of $H^1(\mathcal{U}; \mathbb{R})$ is determined by a Čech cocycle $\{g_{\alpha\beta}\}$, where the $g_{\alpha\beta}$ are locally constant functions on $U_\alpha \cap U_\beta$. We want to produce from this a closed 1-form ω on X, well defined up to the addition of an exact 1-form. If we can find some $\mathscr{C}^\infty$ functions f_α on U_α so that $f_\alpha - f_\beta = g_{\alpha\beta}$ on $U_\alpha \cap U_\beta$, then $df_\alpha = df_\beta$ on the overlaps, so there is a unique 1-form ω on U that is df_α on each U_α. This will be the 1-form we are after. The existence of such functions f_α follows from a general lemma:

Lemma 15.14. *Let $\{f_{\alpha\beta}\}$ be a collection of $\mathscr{C}^\infty$ functions, $f_{\alpha\beta}$ on $U_\alpha \cap U_\beta$,*

satisfying the cocycle conditions: (*i*) $f_{\alpha\alpha} = 0$; (*ii*) $f_{\beta\alpha} = -f_{\alpha\beta}$; *and* (*iii*) $f_{\alpha\gamma} = f_{\alpha\beta} + f_{\beta\gamma}$ *on* $U_\alpha \cap U_\beta \cap U_\gamma$. *Then there are* $\mathscr{C}^\infty$ *functions* f_α *on* U_α, *for all* α, *such that*

$$f_\alpha - f_\beta = f_{\alpha\beta} \qquad \textit{on} \quad U_\alpha \cap U_\beta .$$

Proof. In order to solve these equations $f_\alpha - f_\beta = f_{\alpha\beta}$, we use the existence of a partition of unity subordinate to the covering $\mathcal{U}$ (Appendix B2). This says there are $\mathscr{C}^\infty$ functions φ_α on X, with the closure of the support of φ_α (in X) contained in U_α, with only finitely many φ_α nonzero in a neighborhood of any point, and with $\Sigma_\alpha \varphi_\alpha \equiv 1$. For each α and β define a $\mathscr{C}^\infty$ function $h_{\alpha\beta}$ on U_α by the formula

$$h_{\alpha\beta} = \begin{cases} \varphi_\beta \cdot f_{\alpha\beta} & \text{on } U_\alpha \cap U_\beta, \\ 0 & \text{on } U_\alpha \setminus U_\alpha \cap U_\beta . \end{cases}$$

Now set $f_\alpha = \Sigma_\beta h_{\alpha\beta}$. It is an easy exercise to verify that $h_{\alpha\beta}$ and f_α are $\mathscr{C}^\infty$ functions. To complete the proof, we calculate:

$$\begin{aligned} f_\alpha - f_\beta &= \sum_\gamma (\varphi_\gamma f_{\alpha\gamma} - \varphi_\gamma f_{\beta\gamma}) = \sum_\gamma (\varphi_\gamma(f_{\alpha\gamma} - f_{\beta\gamma}) \\ &= \sum_\gamma \varphi_\gamma(f_{\alpha\beta}) = \Big(\sum_\gamma \varphi_\gamma\Big) f_{\alpha\beta} = 1 \cdot f_{\alpha\beta} = f_{\alpha\beta} . \end{aligned}$$

□

We must verify that the construction made before the lemma gives a well-defined map from $H^1(\mathcal{U};\mathbb{R})$ to H^1X. Suppose first that $\{f_\alpha'\}$ is another collection of functions, with f_α' on U_α such that $f_\alpha' - f_\beta' = g_{\alpha\beta}$ on $U_\alpha \cap U_\beta$, giving rise to the 1-form ω' that is df_α' on U_α. The functions $f_\alpha' - f_\alpha$ on U_α agree on the overlaps, so define a function f on X; and $\omega' - \omega = df$, so ω' and ω define the same element of H^1X. We must also show that if $\{g_{\alpha\beta}'\}$ is a Čech cocycle that is cohomologous to $\{g_{\alpha\beta}\}$, then they determine the same class in H^1X. We are given locally constant functions h_α on U_α such that

$$g_{\alpha\beta}' = -h_\alpha + g_{\alpha\beta} + h_\beta \qquad \text{on} \quad U_\alpha \cap U_\beta .$$

If $f_\alpha - f_\beta = g_{\alpha\beta}$ on $U_\alpha \cap U_\beta$, then $(f_\alpha - h_\alpha) - (f_\beta - h_\beta) = g_{\alpha\beta}'$ on $U_\alpha \cap U_\beta$, and since each h_α is locally constant, $d(f_\alpha - h_\alpha) = df_\alpha$, so they define the same 1-form.

Summarizing what we have so far, we have maps

$$\begin{array}{ccc} H^1X & \longrightarrow & \mathrm{Hom}(H_1X, \mathbb{R}) \\ \uparrow & & \updownarrow \\ H^1(\mathcal{U};\mathbb{R}) & \longleftrightarrow & \{\mathbb{R}\text{-coverings}\}/\cong . \end{array}$$

We want to show that all of these maps are bijections, and that going once clockwise around the diagram, starting at any place, is the identity map on that vector space. We have proved that the top horizontal map in this diagram is one-to-one. It therefore suffices to show that a trip around the diagram, starting with an element ρ in $\mathrm{Hom}(H_1X, \mathbb{R})$, takes ρ to itself. Let $\{f_{\alpha\beta}\}$ be a cocycle for the covering $p_\rho\colon Y_\rho \to X$ defined from this homomorphism. Use Lemma 15.14 to write $f_{\alpha\beta} = f_\alpha - f_\beta$, and let ω be the 1-form that is df_α on U_α. It follows from Proposition 12.22 that H_1X is generated by the classes of closed paths in X. To conclude the proof, it therefore suffices to show that

$$\int_\gamma \omega = \rho([\gamma])$$

when γ is any closed path in X. Both sides depend only on the connected component of X containing the image of γ, so we may assume X is connected. Subdivide the path as in Lemma 15.6, and let γ_i denote the restriction of γ to the ith piece $[t_{i-1}, t_i]$. By Lemma 15.6, $\rho([\gamma]) = \Sigma_{i=1}^n h_i$, where $h_i = f_{\alpha(i)\alpha(i+1)}(x_i)$ for $1 \le i \le n$, with $\alpha(n+1) = \alpha(1)$. Therefore,

$$\begin{aligned}\rho([\gamma]) &= \sum_{i=1}^n f_{\alpha(i)}(x_i) - f_{\alpha(i+1)}(x_i) = \sum_{i=1}^n f_{\alpha(i)}(x_i) - f_{\alpha(i)}(x_{i-1}) \\ &= \sum_{i=1}^n \int_{\gamma_i} df_{\alpha(i)} = \sum_{i=1}^n \int_{\gamma_i} \omega = \int_\gamma \omega .\end{aligned}$$

This completes the proof of Theorem 15.11. □

Corollary 15.15. *There are canonical bijections*

$$H^1X \leftrightarrow \mathrm{Hom}(H_1X, \mathbb{R}) \leftrightarrow \{\mathbb{R}\text{-coverings}\}/\cong \leftrightarrow H^1(\mathcal{U}; \mathbb{R}).$$

Exercise 15.16. Verify directly that the maps of this corollary are linear maps between vector spaces.

The theorem can be used to calculate the De Rham group of more complicated sets in the plane than multiply connected sets, even when it may be difficult to write down differential forms explicitly:

Problem 15.17. (a) Show that if X is the complement of the set $\mathbb{N}$ of nonnegative integers, then H^1X is isomorphic to the space of all infinite sequences $(a_0, a_1, \ldots)$ of real numbers. In particular, H^1X is infinite dimensional. (b) Compute H^1X when X is the complement of the set $(0, 1, 1/2, 1/3, 1/4, \ldots)$. (c) What if X is the complement of any infinite discrete set?

15d. Proof of Mayer–Vietoris for De Rham Cohomology

Now that we have explicitly identified the De Rham cohomology group with the dual of the homology group, we can finish the proof of the Mayer–Vietoris theorem stated in §10d for the De Rham groups. What was missing is the assertion MV(vi) that the map denoted "$-$" from $H^1U \oplus H^1V$ to $H^1(U \cap V)$ is surjective. We also have a linear map

$$H^1U \oplus H^1V \to \mathrm{Hom}(H_1(U) \oplus H_1(V), \mathbb{R})$$

that takes a pair (α, β) to the homomorphism that takes a pair (a, b) in $H_1(U) \oplus H_1(V)$ to $\int_a \alpha + \int_b \beta$. It is a simple exercise to verify, using Theorem 15.11, that this map is also an isomorphism.

The homomorphism "$-$" from $H_1(U \cap V)$ to $H_1U \oplus H_1V$ determines a homomorphism

$$\mathrm{Hom}(H_1(U) \oplus H_1(V), \mathbb{R}) \to \mathrm{Hom}(H_1(U \cap V), \mathbb{R}),$$

and it is a general fact, proved in Appendix C (Lemma C.10), that, since the map that determines it is one-to-one, this map is surjective. Consider the diagram

$$\begin{array}{ccc} H^1U \oplus H^1V & \longrightarrow & H^1(U \cap V) \\ \downarrow & & \downarrow \\ \mathrm{Hom}(H_1(U) \oplus H_1(V), \mathbb{R}) & \longrightarrow & \mathrm{Hom}(H_1(U \cap V), \mathbb{R}) \end{array}$$

where the maps are as just described. We know that each of the vertical maps is an isomorphism, and that the bottom horizontal map is surjective. It follows that the top horizontal map is also surjective, which is the assertion to be proved, provided we check that the diagram commutes. But this is entirely straightforward. Given a class (α, β) in $H^1U \oplus H^1V$, going either way around the diagram, it goes to the homomorphism that takes an element γ in $H_1(U \cap V)$ to the number $\int_\gamma \alpha - \int_\gamma \beta$. □

In fact, the whole Mayer–Vietoris sequence in cohomology is dual to, and can be deduced from, the Mayer–Vietoris sequence in homology. The 0th De Rham group H^0X is dual to the 0th homology group H_0X, as follows. For any function f on X and any 0-cycle $\zeta = \Sigma m_i P_i$ on X, define $f(\zeta) = \Sigma m_i f(P_i)$. If f is locally constant, and ζ is a 0-boundary, then $f(\zeta) = 0$, as follows from the fact that f must take the same values at the two endpoints of any path. This deter-

mines a linear map

$$H^0X \to \operatorname{Hom}(H_0X, \mathbb{R}), \qquad f \mapsto ([\zeta] \mapsto f(\zeta)).$$

Exercise 15.18. Use the fact that H_0X is the free abelian group on the connected components of X to prove that this map from H^0X to $\operatorname{Hom}(H_0X, \mathbb{R})$ is an isomorphism. Show that the reduced groups are also dual: $\widetilde{H}^0X \cong \operatorname{Hom}(\widetilde{H}_0X, \mathbb{R})$.

For simplicity now, for any abelian group A, write A^* for $\operatorname{Hom}(A, \mathbb{R})$. Note that a homomorphism $A \to B$ determines a linear map $B^* \to A^*$. A direct sum decomposition $A = B \oplus C$ determines a direct sum decomposition $A^* = B^* \oplus C^*$. The dual to the Mayer–Vietoris homology sequence is the sequence of vector spaces:

$$0 \longrightarrow H_0(U \cup V)^* \xrightarrow{+} H_0U^* \oplus H_0V^* \xrightarrow{-} H_0(U \cap V)^*$$
$$\xrightarrow{\partial^*} H_1(U \cup V)^* \xrightarrow{+} H_1U^* \oplus H_1V^* \xrightarrow{-} H_1(U \cap V)^* \longrightarrow 0$$

The fact that the homology sequence is exact (Theorem 10.5) implies that this is an exact sequence of vector spaces (see Appendix C2). The isomorphisms from the preceding section can be used to map each term in the Mayer–Vietoris cohomology sequence to the corresponding term in the above sequence:

$$\begin{array}{ccccccccc}
0 \longrightarrow & H^0(U \cup V) & \xrightarrow{+} & H^0U \oplus H^0V & \xrightarrow{-} & H^0(U \cap V) & \xrightarrow{\delta} & H^1(U \cup V) \\
& \downarrow & & \downarrow & & \downarrow & & \downarrow \\
0 \longrightarrow & H_0(U \cup V)^* & \xrightarrow{+} & H_0U^* \oplus H_0V^* & \xrightarrow{-} & H_0(U \cap V)^* & \xrightarrow{\partial^*} & H^1(U \cup V)^*
\end{array}$$

$$\begin{array}{ccccc}
\xrightarrow{+} & H^1U \oplus H^1V & \xrightarrow{-} & H^1(U \cap V) & \longrightarrow 0 \\
& \downarrow & & \downarrow & \\
\xrightarrow{+} & H_1U^* \oplus H_1V^* & \xrightarrow{-} & H_1(U \cap V)^* & \longrightarrow 0.
\end{array}$$

Problem 15.19. (a) Show that each square in the above diagram commutes. (b) Use this to deduce the exactness of the cohomology sequence from the exactness of the homology sequence.

CHAPTER 16

Variations

16a. The Orientation Covering

Every manifold M has a canonical two-sheeted covering $p: \tilde{M} \to M$, called the *orientation covering*, whose fiber over a point P is the two ways to orient M at P. Čech cocycles provide a convenient way to construct this covering. Let $G = \{\pm 1\}$ be the group of order two, and take an open covering $\mathcal{U} = \{U_\alpha\}$ of M to be images of the coordinate charts $\varphi_\alpha: V_\alpha \to U_\alpha \subset M$, with V_α open in $\mathbb{R}^n$. The Jacobian determinant of the change of coordinates from V_α to V_β has a locally constant sign, which gives a locally constant function from $U_\alpha \cap U_\beta$ to $\{\pm 1\}$. The chain rule for Jacobians implies that this is a cocycle. Define $p: \tilde{M} \to M$ to be the resulting $\{\pm 1\}$-covering. An *orientation* of M can be defined to be a section of this covering, i.e., a continuous mapping $\sigma: M \to \tilde{M}$ such that $p \circ \sigma$ is the identity map on M.

Exercise 16.1. (a) Show that this orientation covering is independent of the choice of coordinate charts. (b) If M is connected, show that $\tilde{M}$ has one or two connected components: one if M is nonorientable, two if M is orientable. (c) Show that $\tilde{M}$ is always orientable.

Exercise 16.2. When $M = \mathbb{RP}^2$ show that the orientation covering is isomorphic to the covering $S^2 \to \mathbb{RP}^2$. Do the same for $\mathbb{RP}^n$, $n > 2$. Identify the orientation covering for the Klein bottle.

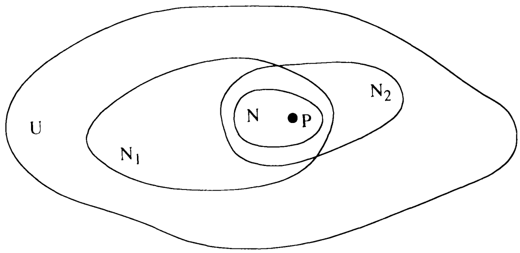
16b. Coverings from 1-Forms

From Corollary 15.7 and Theorem 15.11 we have a canonical bijection

$$H^1X \leftrightarrow \{\mathbb{R}\text{-coverings of } X\}/\cong.$$

It follows that there is an $\mathbb{R}$-covering $p_\omega\colon X_\omega \to X$ of an open set X in the plane corresponding to a closed 1-form ω on X, with two 1-forms giving isomorphic coverings when their difference is exact. To construct this covering, following the procedure given in §15c, one takes a collection $\mathcal{U}$ of open sets U_α on each of which $\omega = df_\alpha$ for some function f_α, and then uses the transition functions $g_{\alpha\beta} = f_\alpha - f_\beta$ to patch together trivial $\mathbb{R}$-coverings of each U_α into a global $\mathbb{R}$-covering of X. In this section we describe a more direct way to construct this covering.

The idea is to put all graphs of functions f on open subsets of X such that $df = \omega$ together into one big covering space. A good language for describing this is that of germs. A *germ* of a $\mathscr{C}^\infty$ function *at* a point P in X is an equivalence class of $\mathscr{C}^\infty$ functions defined in neighborhoods of P. Two functions f_1 on a neighborhood N_1 and f_2 on N_2 are defined to be equivalent if there is a neighborhood N of P contained in $N_1 \cap N_2$ such that f_1 and f_2 are equal on N.



These germs can be added, multiplied, and differentiated, just like functions on definite open sets.

Exercise 16.3. Verify that this is an equivalence relation, and that these operations preserve equivalence classes.

A germ at P has a value at P (the value $f(P)$ of any representative f), but it does not have a value at any other point. The idea is simply that we only care about functions near the point P, and we allow ourselves to shrink the neighborhoods arbitrarily. Given a function f

on a neighborhood of P, the equivalence class containing it is called the germ *defined by* f *at* P.

Since derivatives of germs make sense, it makes sense to say that the *differential* of a germ at a point is a given 1-form; if the 1-form is $\omega = p\,dx + q\,dy$, the differential of a germ of f at P is ω if

$$\frac{\partial f}{\partial x}(P) = p(P) \quad \text{and} \quad \frac{\partial f}{\partial y}(P) = q(P).$$

If ω is a closed 1-form on X, define X_ω to be the set of all germs of functions at points of X whose differential is ω. We will make this into a topological space, so that the map $p_\omega: X_\omega \to X$, that takes a germ to the point at which it is a germ, is a covering map. For each open set N in X and function f on N such that $df = \omega$ on N, define a basic open set N_f in X_ω:

$$N_f = \{\text{germs at points of } N \text{ defined by } f\}.$$

A set in X_ω is defined to be open if it is a union of basic sets N_f.

If N is an open set in X such that ω is exact on N (for example, any open disk or rectangle in X), then $p_\omega{}^{-1}(N)$ is a disjoint union of the open sets N_f, where f runs through all functions on N whose differential is ω. The point is that if f is one such function, then all others, on N or on any open subset of N, are of the form $f + c$ for some locally constant function c. This shows that the covering is trivial over N, and, since any point in X has such neighborhoods, this shows that we have a covering space:

$$p_\omega: X_\omega \to X.$$

In addition, this is an $\mathbb{R}$-covering. The action of a real number on X_ω takes a germ to the sum of the germ and the real number. The preceding discussion shows that this is an even action of $\mathbb{R}$ on X_ω, and that the covering is an $\mathbb{R}$-covering.

Exercise 16.4. Verify that when $X = \mathbb{R}^2 \setminus \{(0,0)\}$ and $\omega = d\vartheta$, then the polar coordinate covering of X is a connected component of X_ω.

Choosing any function f on any connected open subset N of X with $df = \omega$ on N determines a connected component of X_ω. Namely, this is the connected component containing the open set N_f.

Exercise 16.5. Show that this component is the union of all open sets of the form $N'_{f'}$ with N' a connected open set in X and $df' = \omega$ on N' such that there is a chain of connected open sets $N = N_0$,

$N_1, \ldots, N_r = N'$ in X, with functions f_i on N_i such that $df_i = \omega$, and $f_i = f_{i+1}$ on $N_i \cap N_{i+1}$.

In general, there is a natural bijection between X_ω and the product $X \times \mathbb{R}$, that takes a germ at P to the pair $P \times r$, where r is the value of the germ at P. If $\mathbb{R}$ is has its usual topology, this bijection is continuous, but it is *not* a homeomorphism. In the polar coordinate example, each connected component of X_ω maps to a closed submanifold of the product $X \times \mathbb{R}$; these are the graphs of the multivalued functions "$\vartheta + c$." In general, however, this need not be the case. Problem 6.22 shows that, when X is the complement of two points P and Q, and ω is a linear combination $r\, d\vartheta_P + s\, d\vartheta_P$ with r/s irrational, the sheets of a component of X_ω can come arbitrarily close together, if regarded as a subset of $X \times \mathbb{R}$. With this understanding, the connected components of X_ω can be regarded as the multivalued functions on X whose differential is ω—but their graphs need not be closed in $X \times \mathbb{R}$.

Problem 16.6. (a) Show that this covering $p_\omega \colon X_\omega \to X$ is isomorphic to the covering constructed with cocycles. (b) Show directly that 1-forms that differ by an exact form determine isomorphic coverings.

Problem 16.7. Assume X is connected, and let Y be a connected component of X_ω. Show that $Y \to X$ is a G-covering, where G is the period module, i.e., G is the subgroup of $\mathbb{R}$ generated by the periods of ω; these periods are the numbers $\int_\gamma \omega$, as γ varies over loops at x (or equivalently, closed 1-chains on X).

16c. Another Cohomology Group

In addition to the De Rham and Čech cohomology groups, there are cohomology groups $H^0(X; G)$ and $H^1(X; G)$ that one can assign to any topological space X, where G can be any abelian group. Lacking smooth functions and 1-forms to evaluate on points and paths, one makes these groups directly out of objects that are defined by their values on points and paths. Define a *0-cochain* to be an arbitrary function from X to G. Define a *1-cochain* to be a function from the set of all continuous paths on X to G, with the requirement that the function takes all constant paths to the identity element 0 in G. Note that, unlike the case of 0-chains or 1-chains, cochains can assign nonzero

elements of G to arbitrarily many elements. Since the set of functions from any set to an abelian group G is an abelian group (by adding the values in G), the 0-cochains form a group denoted $C^0(X;G)$, and the 1-cochains form a group $C^1(X;G)$.

If $a\colon X \to G$ is a 0-cochain, define the *coboundary* δa to be the 1-cochain whose value on a path γ is $a(\gamma(1)) - a(\gamma(0))$. Define the 0th *cohomology group* $H^0(X;G)$ be the set of 0-cochains whose boundary is zero.

Exercise 16.8. Show that giving an element of $H^0(X;G)$ is equivalent to giving a function from the set of path-connected components of X to G. Deduce an isomorphism $H^0(X;G) \cong \text{Hom}(H_0X, G)$.

If c is a 1-cochain, it defines a homomorphism from the 1-chains C_1X to G by the formula $c(\Sigma n_i\gamma_i) = \Sigma n_i c(\gamma_i)$. A 1-cochain c is called a 1-*cocycle* if $c(\partial\Gamma) = 0$ for all continuous $\Gamma\colon [0,1] \times [0,1] \to X$. The 1-cocycles form a subgroup $Z^1(X;G)$ of $C^1(X;G)$. Every coboundary δa is a 1-cocycle, so the 1-coboundaries form a subgroup $B^1(X;G)$ of $Z^1(X;G)$. The quotient group is the *first cohomology group*

$$H^1(X;G) \;=\; Z^1(X;G)/B^1(X;G).$$

There is a natural homomorphism from $H^1(X;G)$ to $\text{Hom}(X_1X, G)$, that takes the class of a 1-cocycle c to the homomorphism that takes the homology class of a 1-cycle γ to the element $c(\gamma)$.

Exercise 16.9. Verify that this is a well-defined homomorphism.

Proposition 16.10. *The homomorphism $H^1(X;G) \to Hom(H_1X, G)$ is an isomorphism.*

Proof. To give a homomorphism from $H_1X = Z_1X/B_1X$ to G is equivalent to giving a homomorphism from Z_1X to G that vanishes on B_1X; that is,

$$\text{Hom}(H_1X, G) \;=\; \text{Kernel}(\text{Hom}(Z_1X, G) \to \text{Hom}(B_1X, G)).$$

We have homomorphisms

$$B^1(X;G) \;\to\; Z^1(X;G) \;\to\; \text{Hom}(Z_1X, G) \;\to\; \text{Hom}(B_1X, G),$$

where the middle map takes the cocycle c to the homomorphism $\gamma \to c(\gamma)$. The assertion of the proposition is easily seen to be equivalent to the assertion that this sequence is exact at the two middle groups.

The fact that the image of each map is contained in the next is the

content of the preceding exercise. To prove the opposite inclusions, we need to choose arbitrarily a point x_α in each path component X_α of X, and choose for each point y in X an arbitrary path τ_y starting at the chosen point x_α of the component containing y and ending at y.

Suppose c in $Z^1(X;G)$ maps to the zero homomorphism from Z_1X to G. We want to construct a 0-chain whose coboundary is c. Define the 0-chain a by the formula $a(y) = c(\tau_y)$. We claim that $(\delta a)(\gamma) = c(\gamma)$ for all paths γ. In fact,

$$(\delta a)(\gamma) = a(\gamma(1)) - a(\gamma(0)) = c(\tau_{\gamma(1)}) - c(\tau_{\gamma(0)}),$$

and $c(\tau_{\gamma(1)}) - c(\tau_{\gamma(0)}) = c(\gamma)$ since $\tau_{\gamma(1)} - \tau_{\gamma(0)} - \gamma$ is a 1-cycle, and c is assumed to vanish on 1-cycles.

To finish the proof, we must show that if $f\colon Z_1X \to G$ is a homomorphism that vanishes on B_1X, then f comes from some 1-cocycle c. Define the cochain c by the formula

$$c(\gamma) = f(\gamma + \tau_{\gamma(0)} - \tau_{\gamma(1)})$$

for any path γ. Equivalently, for any 1-chain γ, $c(\gamma) = f(\gamma - \Sigma m_i \tau_{y_i})$, where $\Sigma m_i y_i = \partial(\gamma)$. In particular, if γ is a 1-cycle, $c(\gamma) = f(\gamma)$, and if $\gamma = \partial\Gamma$, $c(\partial\Gamma) = f(\partial\Gamma) = 0$. So c is a 1-cocycle that maps to f. □

This proposition implies that $H^1(X;G)$ is isomorphic to the Čech group $H^1(\mathcal{U};G)$, provided $\mathcal{U}$ is a suitable cover of a nice space, so that Corollary 15.7 applies.

Exercise 16.11. When X is an open set in the plane, and $G = \mathbb{R}$, construct homomorphisms $H^0X \to H^0(X;\mathbb{R})$ and $H^1X \to H^1(X;\mathbb{R})$ from the De Rham groups to these cohomology groups, and show that they are isomorphisms.

There is also a Mayer–Vietoris theorem for these cohomology groups. If U' is an open subset of U there are natural restriction maps from $H^0(U;G)$ to $H^0(U';G)$ and from $H^1(U;G)$ to $H^1(U';G)$. If X is a union of two open sets U and V, there is a coboundary map

$$\delta\colon H^0(U \cap V;G) \to H^1(U \cup V;G).$$

To define this, given a 0-cocycle a on $U \cap V$, extend a to a 0-cochain $\tilde{a}$ on all of $U \cup V$ by defining $\tilde{a}$ to be zero on all points not in $U \cap V$. Then define the 1-cocycle $\delta(a)$ on $U \cup V$ whose value on a path γ is obtained by writing $\gamma = \gamma_1 + \gamma_2 + \tau$, where γ_1 is a 1-chain on U, γ_2 a 1-chain on V, and τ is a 1-boundary (see the proof of Lemma 10.2), and setting $\delta(a)(\gamma) = \tilde{a}(\partial\gamma_1)$.

Problem 16.12. Show that this definition is independent of choices, and verify that the resulting Mayer–Vietoris sequence

$$0 \longrightarrow H^0(U\cup V;G) \xrightarrow{+} H^0(U;G)\oplus H^0(V;G) \xrightarrow{-} H^0(U\cap V;G)$$

$$\xrightarrow{\delta} H^1(U\cup V;G) \xrightarrow{+} H^1(U;G)\oplus H^1(V;G) \xrightarrow{-} H^1(U\cap V;G)$$

is exact.

16d. G-Sets and Coverings

In this section and the next we describe two general constructions, which may help to put the constructions of §14a in context. The exercises verify the assertions made about these constructions.

Let $p\colon Y \to X = Y/G$ be a G-covering, without base point for the moment. Suppose T is any set, and we have a left action of G on T; we say that T is a *G-set*. Give T the discrete topology, so an action of G on T is a mapping from $G\times T$ to T satisfying properties (1) and (2) of §11c. The group G acts on the left on $Y\times T$ by the formula $g\cdot(y\times t) = g\cdot y\times g\cdot t$. Define Y_T to be the space of orbits:

$$Y_T = (Y\times T)/G.$$

Write $\langle y\times t\rangle$ in Y_T for the orbit containing $y\times t$. Let $p_T\colon Y_T\to X$ be the mapping that sends $\langle y\times t\rangle$ to $p(y)$.

Exercise 16.13. Show that the mapping $p_T\colon Y_T\to X$ is a covering map.

If T and T' are sets with G-actions, a *map* of G-sets is a function $\varphi\colon T\to T'$ such that $\varphi(g\cdot t) = g\cdot\varphi(t)$ for all t in T and g in G. A map of G-sets is an *isomorphism* if it is bijective, so there is an inverse mapping of G-sets from T' to T. A map $\varphi\colon T\to T'$ of G-sets determines a continuous mapping from $Y\times T$ to $Y\times T'$, taking $y\times t$ to $y\times\varphi(t)$. This is compatible with the actions of G, so it determines a continuous mapping from Y_T to $Y_{T'}$, taking $\langle y\times t\rangle$ to $\langle y\times\varphi(t)\rangle$, which commutes with the projections to X. If φ is an isomorphism, this is an isomorphism of coverings. Conversely, we have:

Exercise 16.14. Let $p\colon Y\to X = Y/G$ be a G-covering, with Y connected. Show that the two G-sets determine isomorphic coverings of X if and only if the G-sets are isomorphic. Show in fact that any continuous mapping f from Y_T to $Y_{T'}$ commuting with the projections to X comes from a map of G-sets from T to T'.

Exercise 16.15. If $T \to T'$ and $T' \to T''$ are maps of G-sets, show that the composite of $Y_T \to Y_{T'}$ and $Y_{T'} \to Y_{T''}$ is the mapping determined by the composite $T \to T''$.

Suppose the action of G on a set T is *transitive*: for any t_1 and t_2 in T, there is some g in G such that $g \cdot t_1 = t_2$. If a point t is chosen, let $H \subset G$ be the subgroup of elements of G that fix t, i.e., $H = \{g \in G: g \cdot t = t\}$. Then the G-set G/H of left cosets is isomorphic to the G-set T, by mapping the coset gH containing g to the point $g \cdot t$.

Exercise 16.16. Show that two transitive G-sets are isomorphic if and only if the corresponding subgroups of G are conjugate.

In Exercise 11.27 we saw how a subgroup H of G determines a covering $Y/H \to X$.

Exercise 16.17. If $T = G/H$ is a set of left cosets, show that the covering $Y_T \to X$ is isomorphic to the covering $Y/H \to X$.

Exercise 16.18. If T is a disjoint union of G-sets T_α, show that the covering $Y_T \to X$ is a disjoint union of the coverings $Y_{T_\alpha} \to X$.

An arbitrary G-set T is a disjoint union of its orbits T_α. By the preceding exercises the corresponding covering $Y_T \to X$ is a disjoint union of coverings of the form Y/H_α, for H_α subgroups of G.

Now suppose X is connected and locally path-connected, and has a universal covering space $\tilde{X} \to X$. Choose a point $\tilde{x}$ of $\tilde{X}$ lying over x in X. We have seen that, with these choices, the universal covering is a $\pi_1(X,x)$-covering. So every action of the fundamental group on a set T determines a covering $\tilde{X}_T \to X$. Combining what we have just proved with Proposition 13.23, we have:

Proposition 16.19. *Every covering of X is isomorphic to one obtained from the universal covering $\tilde{X} \to X$ by a left action of $\pi_1(X,x)$ on some set T. This covering is connected if and only if the action on T is transitive. Two such coverings are isomorphic if and only if the $\pi_1(X,x)$-sets are isomorphic.*

A left action of a group G on a set T is the same as a homomorphism of G to the group Aut(T) of permutations of T. Two such homomorphisms give isomorphic G-sets exactly when the homo-

morphisms are conjugate. In particular, taking $G = \pi_1(X, x)$ and $T = \{1, \ldots, n\}$, so Aut(T) is the symmetric group $\mathfrak{S}_n$ on n letters, we have a canonical bijection

$$\{n\text{-sheeted coverings of } X\}/\cong \;\leftrightarrow\; \mathrm{Hom}(\pi_1(X,x), \mathfrak{S}_n)/\text{conjugacy}.$$

16e. Coverings and Group Homomorphisms

Suppose $p\colon Y \to X$ is a G-covering, and $\psi\colon G \to G'$ is any homomorphism of groups. If we make G' into a left G-set, the construction of the preceding section can be used to construct a covering of X that is locally a product $X \times G' \to X$. We want to do this in such a way that this covering can be made into a G'-covering. The simplest way to make G' into a left G-set is by defining, for g in G and g' in G', $g \cdot g'$ to be $\psi(g) \cdot g'$, the latter using the group product in G'. However, we will use another, which will allow us to define a compatible left action of G' on the result. Define the left action of G on G' by

$$g \cdot g' = g' \cdot \psi(g^{-1}) = g' \cdot \psi(g)^{-1} \qquad \text{for } g \in G \text{ and } g' \in G'.$$

It is straightforward to check that this is a left action of G on G'. Now define $p(\psi)\colon Y(\psi) \to X$ to be the covering constructed from this left G-set. That is, $Y(\psi)$ is the quotient of $Y \times G'$ by the left action of G by $g \cdot (y \times g') = (g \cdot y \times g' \cdot \psi(g^{-1})$, with projection from $Y(\psi)$ to X determined by p on the first factor. Make G' act on $Y(\psi)$ by $g' \cdot \langle y \times h' \rangle = \langle y \times g' \cdot h' \rangle$.

Exercise 16.20. Show that $Y(\psi) \to X$ is a G'-covering of X. If Y has a base point y, $Y(\psi)$ gets the base point $\langle y \times e \rangle$. When p is the universal covering of X, this is the construction of Proposition 14.1.

Exercise 16.21. If $Y_\rho \to X$ is the G-covering determined by a homomorphism $\rho\colon \pi_1(X, x) \to G$, and $\psi\colon G \to G'$ is a homomorphism, show that the G'-covering $Y_\rho(\psi) \to X$ constructed from $Y_\rho \to X$ and ψ is isomorphic to the G'-covering $Y_{\psi\circ\rho} \to X$.

Exercise 16.22. Let $p\colon Y \to X$ be a G-covering, with Y connected. Let $\psi_1\colon G \to G'$ and $\psi_2\colon G \to G'$ be two homomorphisms. Show that, if base points are taken into account, the coverings $Y(\psi_1)$ and $Y(\psi_2)$ are isomorphic if and only if ψ_1 and ψ_2 are equal. Without base points, they are isomorphic if and only if ψ_1 and ψ_2 are conjugate, i.e., there is some g' in G' such that $\psi_2(g) = g' \cdot \psi_1(g) \cdot (g')^{-1}$ for all g in G.

Exercise 16.23. Show that if G is abelian, any G-covering of X is obtained from the universal abelian covering $\tilde{X}_{\text{abel}} \to X$ by a unique homomorphism $\psi: H_1 X \to G$.

16f. G-Coverings and Cocycles

In §16d we constructed from a G-covering $p: Y \to X$ and a left action of G on a set T, a new covering $p_T: Y_T \to X$, where Y_T is the quotient of $(Y \times T)/G$, with G acting simultaneously on the left on Y and on T. This can also be described by a cocycle. Take a cocycle $\{g_{\alpha\beta}\}$ of transition functions from a trivialization of the covering $Y \to X$ over $\mathcal{U} = \{U_\alpha\}$, and use them to glue together the disjoint union of the $U_\alpha \times T$, identifying $x \times t$ in $U_\alpha \times T$ with $x \times g_{\beta\alpha}(x) \cdot t$ in $U_\beta \times T$ if x is in $U_\alpha \cap U_\beta$.

Exercise 16.24. Verify that this covering defined by a cocycle is isomorphic to that defined in §16d.

If $\psi: G \to G'$ is a homomorphism of groups, then ψ determines a map

$$\psi_*: H^1(\mathcal{U}; G) \to H^1(\mathcal{U}; G')$$

that takes a class represented by a cocycle $\{g_{\alpha\beta}\}$ to the class of $\{\psi(g_{\alpha\beta})\}$. (Exercise: Verify that this is well defined.) This means that any G-covering of X that is trivial over each U_α determines a G'-covering of X, also trivial over each U_α: if the G-covering is constructed by gluing the products $U_\alpha \times G$ with the transition functions $x \times g \mapsto x \times g \cdot g_{\alpha\beta}(x)$, then the corresponding G'-covering is constructed by gluing the products $U_\alpha \times G'$ with the transition functions $x \times g' \mapsto x \times g' \cdot \psi(g_{\alpha\beta}(x))$.

Problem 16.25. Show that this construction of a G'-covering from a G-covering agrees with the construction in §16d as the quotient of $Y \times G'$ by the action of G: $g \cdot (y \times g') = g \cdot y \times g' \cdot \psi(g^{-1})$: In particular, it is independent of choice of $\mathcal{U}$.

Exercise 16.26. A homomorphism $\psi: G \to G'$ determines a map from $\text{Hom}(\pi_1(X,x), G)$ to $\text{Hom}(\pi_1(X,x), G')$, taking ρ to $\psi \circ \rho$. Show that this is compatible with the bijections of Proposition 14.1.

In our definition of Čech cohomology $H^1(\mathcal{U}; G)$ there has been no topology on the group G, or, more accurately, we have equipped G with the discrete topology. Čech cohomology is also important when G is a group with a more interesting topology. One can define a set $H^1(\mathcal{U}; G)$ the same way, requiring the functions $g_{\alpha\beta}$ and h_α in the definition to be continuous maps from $U_{\alpha\beta}$ and U_α to G, respectively. From a cocycle $\{f_{\alpha\beta}\}$ one constructs in the same way a (principal) G-*bundle*, i.e., a space Y, on which G acts continuously, with a mapping $p: Y \to X$ so that $p(g \cdot y) = p(y)$, locally isomorphic to the projection $X \times G \to X$ (but in this product G is a topological space, not necessarily discrete).

If the topological group G acts continuously on the left on a space V, then a G-bundle determines a bundle with fiber V. The construction is as in the discrete case: take trivial coverings $U_\alpha \times V$, and glue them together using the identifications from $U_\alpha \times V$ to $U_\beta \times V$ over $U_\alpha \cap U_\beta$ by $x \times v \mapsto x \times g_{\beta\alpha}(x) \cdot v$; or one may construct the quotient $(Y \times V)/G$. The fundamental example is the case $G = \mathrm{GL}_n\mathbb{R}$ (with its usual topology as an open subset of $\mathbb{R}^{n^2}$), with $V = \mathbb{R}^n$, and the resulting bundle is called a *vector bundle* of rank n.

Problem 16.27. For $n = 1$, and $\mathcal{U}$ a covering of a plane open set U as above, show that the homomorphism from $H^1(\mathcal{U}; \{\pm 1\})$ to $H^1(\mathcal{U}; \mathrm{GL}_1\mathbb{R})$ determined by the inclusion of $\{\pm 1\}$ in $\mathrm{GL}_1\mathbb{R}$ is an isomorphism. Equivalently, giving a line bundle is equivalent to giving a two-sheeted covering. Prove this directly. Compute this group for U an annulus. Over a circle in the annulus you should find two line bundles, corresponding to a cylinder and a Moebius band.

Problem 16.28. Show how to make the set of all tangent vectors to the sphere $S^2 \subset \mathbb{R}^3$ into a vector bundle of rank 2. Generalize to construct the tangent bundle of an arbitrary manifold.

PART IX
TOPOLOGY OF SURFACES

The goal is to extend what we have done for open sets in the plane to open sets in surfaces, especially compact orientable surfaces. Some of this is straightforward generalization, valid for any surface, given basic notions about coordinate charts, but there are some new features that are special to the compact and orientable case.

We show that every compact orientable surface is homeomorphic to a sphere with g handles, at least under the assumption that the surface can be triangulated. Any such surface can be realized by taking a convex polygon with $4g$ sides, and making suitable identifications on the boundary. From this description it is routine to compute the fundamental group and first homology group.

In Chapter 18 we show that for any differentiable surface X, there is a canonical isomorphism between the De Rham group H^1X of closed mod exact 1-forms, and the dual group $\mathrm{Hom}(H_1X, \mathbb{R})$ to the homology group. This leads to a definition and calculation of an intersection pairing on the homology group H_1X.

Note: A surface is assumed to be connected unless otherwise stated. See Appendix D for foundational facts about surfaces.

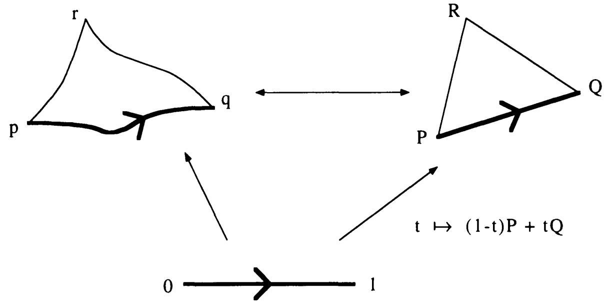
CHAPTER 17

The Topology of Surfaces

17a. Triangulation and Polygons with Sides Identified

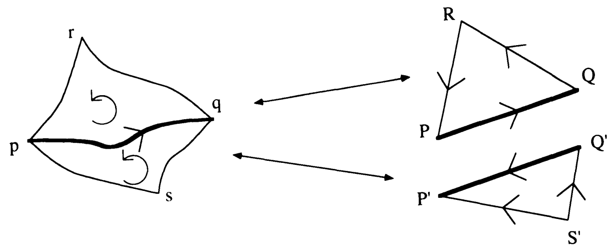
Our aim in this section is to show that a compact orientable surface is homeomorphic to a sphere with handles, under the assumption that the surface can be triangulated. The idea is to "flatten" it out, to realize the surface as a polygon in the plane, with certain identifications on the boundary edges, and then, by cutting and pasting, to simplify these realizations until we can recognize the result.

A *triangulation* of a compact surface X has a finite number f of *faces*, each of which is a subset of X homeomorphic to an ordinary (closed) triangle in the plane, with three *edges* homeomorphic to closed intervals, and three *vertices* that are points. Two edges can meet only at a vertex, and two faces meet only at one vertex or exactly along a common edge; in the latter case the faces are called *adjacent*. There are many choices for such homeomorphisms between plane triangles and faces of the triangulation. In order to be able to compare the homeomorphisms on adjacent faces in X, we first choose a homeomorphism of each edge on X with the interval $[0, 1]$. We can then find homeomorphisms of each face on X with a plane triangle so that, on the edges, the maps are those determined by these homeomorphisms:

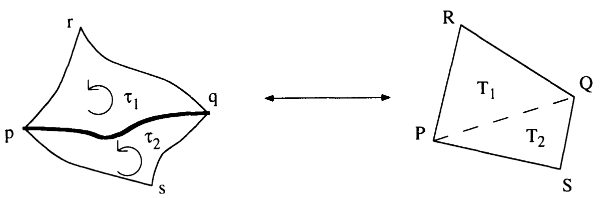


If two triangles correspond to adjacent faces, this means that the resulting homeomorphism between corresponding sides $\overline{PQ}$ and $\overline{P'Q'}$ is given by the "affine" map $(1-t)P + tQ \mapsto (1-t)P' + tQ'$. When we identify sides of polygons, it will always be by such affine homeomorphisms, so there will be no ambiguity in the results.

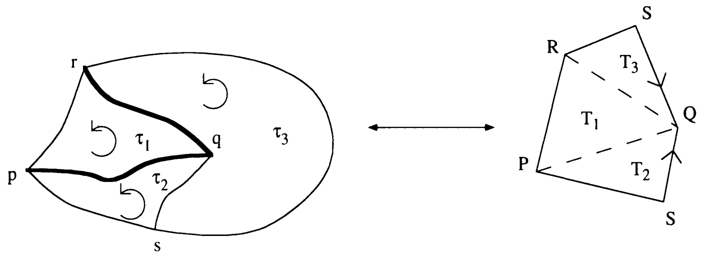
An orientation of X determines a counterclockwise direction around the boundary of each triangle, with adjacent triangles determining opposite directions along their common edge:



Choose one face τ_1 in the triangulation, and choose a homeomorphism of τ_1 with a plane triangle $\Pi_1 = T_1$, with counterclockwise orientation. Choose a face τ_2 adjacent to τ_1, and extend the homeomorphism along their common edge to a homeomorphism of τ_2 with a triangle T_2 adjacent to T_1:

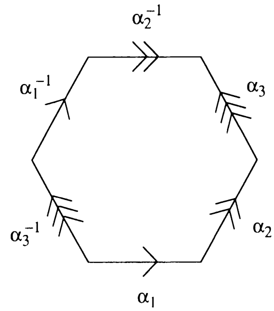


Choose T_2 so the union of T_1 and T_2 is a convex quadrilateral Π_2. This gives a homeomorphism from $\tau_1 \cup \tau_2$ to Π_2. Now choose τ_3 adjacent to τ_1 or τ_2 (if τ_3 is adjacent to both, choose one of them arbitrarily), and extend the map from the common edge of τ_3 and the (chosen) τ_1 or τ_2 to a homeomorphism from τ_3 to a triangle T_3 adjacent to the corresponding side of Π_2. Let Π_3 be the resulting five-sided polygon, which we can take to be convex:



The arrows indicate the identification made in the map from the figure on the right onto that on the left. Continue in this way, until all f of the faces have been used. We then have a convex polygon $\Pi_f = \Pi$ with $f + 2$ sides. Each side of Π will correspond to a common edge of two faces on X, and each edge occurring this way will correspond to two such sides of Π. Thus the sides of Π will be paired off, and we have a continuous map from Π to X that realizes X as a quotient space of Π, obtained by identifying corresponding points on corresponding sides. Note that, when traveling around Π in a counterclockwise direction, one travels along two corresponding sides in opposite directions.

This realization of X as an identification space of a polygon is half way to our goal. The data used in constructing X from the polygon can be encoded by taking an alphabet with m letters, with $m = \frac{1}{2}(f + 2)$, and writing down a sequence of $2m$ symbols, each of which is a letter α in the alphabet, or its "inverse" α^{-1}, with each of these symbols occurring just once. For example, the space constructed by making the identifications indicated in the following diagram



can be described by the code $\alpha_1\alpha_2\alpha_3\alpha_2^{-1}\alpha_1^{-1}\alpha_3^{-1}$.

Exercise 17.1. Show that this code represents a torus.

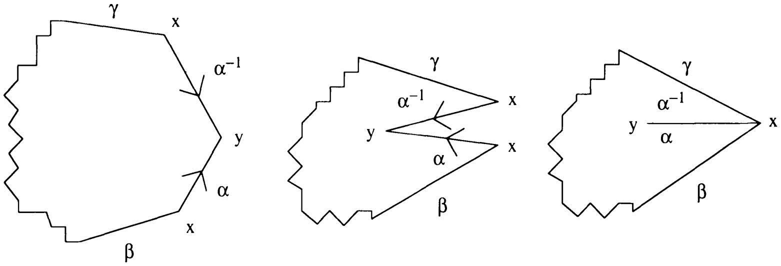
Exercise 17.2. If the number of sides is four, show that X is homeomorphic to a sphere or a torus.

There is a good deal of arbitrariness in the choice of polygon and its code, even for a given triangulation and ordered choice of faces of X. The polygon can certainly be replaced by any other convex polygon with the same number of sides. (For simplicity, one may use a regular polygon.) Of course, even when the polygon and the identifications are fixed, the choice of alphabet is arbitrary, as is the choice of which of a pair of corresponding sides is written as a letter α of the alphabet and which as α^{-1}. In addition, the place where one starts listing the sides is arbitrary, and the code can therefore by cyclically permuted. Note that the identifications are made by identifying corresponding points on corresponding sides; some vertices of Π may become identified by this, and others may not.

17b. Classification of Compact Oriented Surfaces

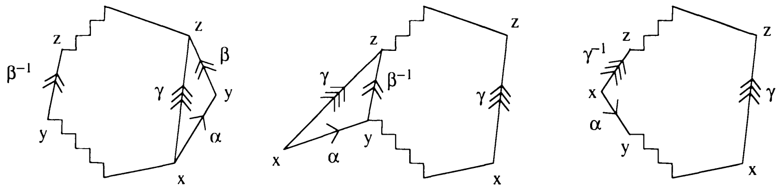
For the rest of the argument we do not need the triangulation, but only this description as a polygon with sides identified according to a code. We want to prescribe some rules for simplifying a polygonal presentation of a surface. We proceed in several steps. By Exercise 17.2 we can assume the polygon has $2m$ sides, with $m \geq 3$.

Step 1. If a letter α and its inverse α^{-1} occur successively in the code, they can both be omitted. That is, X is homeomorphic to the surface obtained from a convex polygon with $2m-2$ sides, using the same code in the same order but with α and α^{-1} omitted. This can be seen from the pictures



The point is that if one first carries out the identification of α with α^{-1}, one gets a space homeomorphic to a convex polygon, so that the remaining identifications are prescribed by the code with α and α^{-1} omitted.

Step 2. It can be assumed that all the vertices of the polygon map to the same point of X. To show this, it suffices, if not all the vertices have the same image, to show how to increase by one the number of vertices mapping to a given point x, without changing the total number of vertices. There is a side, say labeled α, that joins two vertices, one of which is mapped to x and the other to a point y not equal to x. Let β be the side of Π adjacent to α at the point mapping to y. Draw the diagonal γ as shown on the polygon, and cut off the triangle formed by α, β, and γ, and attach it to the polygon by identifying β with β^{-1}:



The surface X is the identification space of this new polygon, and it has the same number of sides, but one more vertex mapping to x.

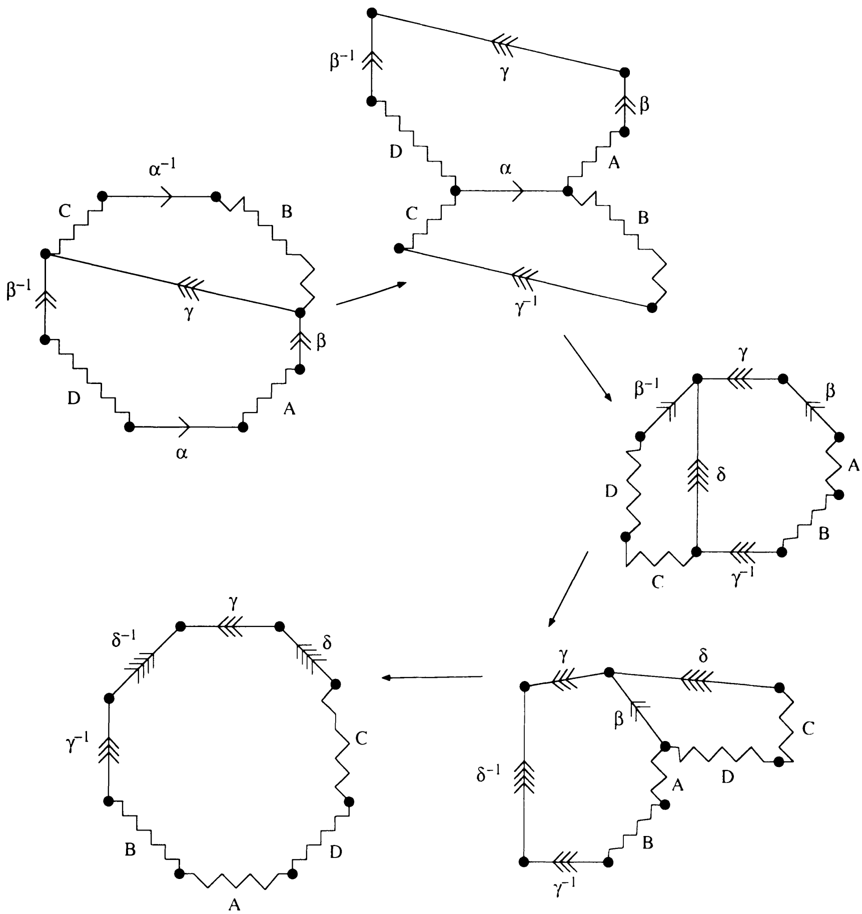
Step 3. By steps 1 and 2 we may assume that all vertices map to the same point x in X, and that no edge is adjacent to its inverse. We claim next that for any edge α there must be an edge β lying between α and α^{-1} so that β^{-1} lies between α^{-1} and α; that is, in a code, after cyclically permuting if necessary, they occur in the order

$$\ldots \alpha \ldots \beta \ldots \alpha^{-1} \ldots \beta^{-1} \ldots \quad .$$

If not, we could construct the identification space X by first identi-

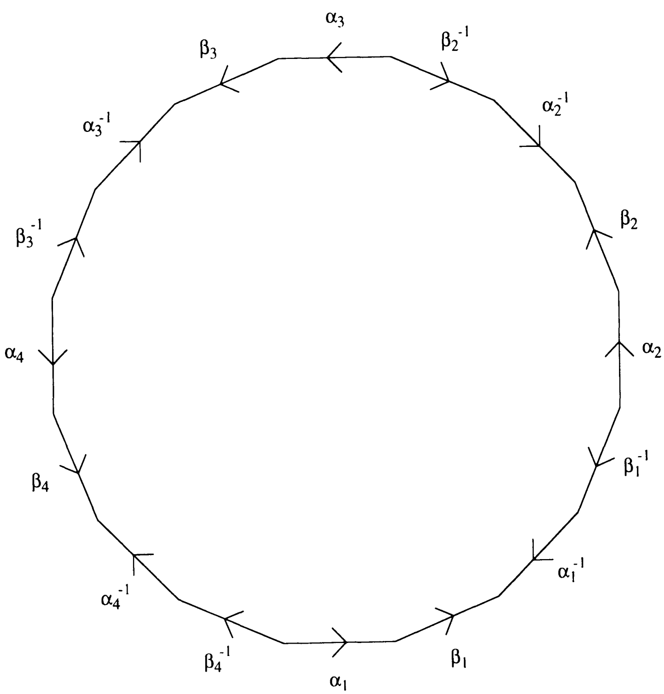
fying α with α^{-1}, and then separately doing all the identification prescribed by edges lying (counterclockwise) between α and α^{-1}, and those prescribed by edges lying (counterclockwise) between α^{-1} and α. The two endpoints of α never get identified in this process, contradicting the assumption that all the vertices map to the same point of X.

Now we simplify the code as follows: Choose an edge α, and take an edge β so that the edges α, β, α^{-1}, and β^{-1} occur in this order, but perhaps with other edges in between some of these (denoted A, B, C, and D in the following diagram). Perform the following sequence of cutting and pasting moves:

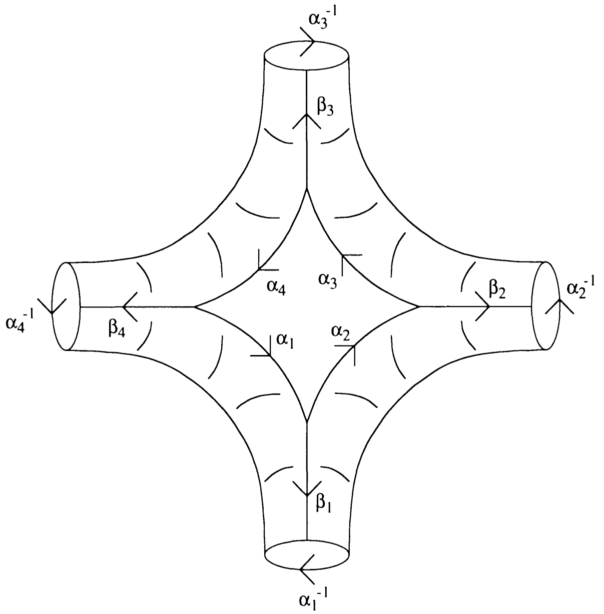


This gives a new polygon with the same properties and the same number of vertices, which also represents X, but now it has a successive sequence $\delta \cdot \gamma \cdot \delta^{-1} \cdot \gamma^{-1}$. In addition, if other such sequences occurred elsewhere in the polygon (i.e., in one of the parts labeled A, B, C, or D), such sequences are not disturbed by this procedure. So we may continue this process, until we have represented X as a polygon with sides identified according to a code

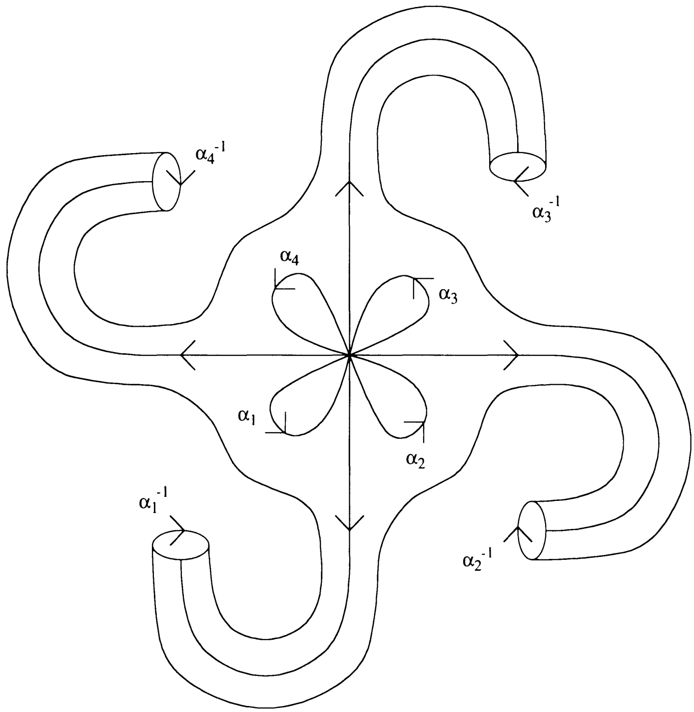
(17.3)
$$\alpha_1 \cdot \beta_1 \cdot \alpha_1^{-1} \cdot \beta_1^{-1} \cdot \alpha_2 \cdot \beta_2 \cdot \alpha_2^{-1} \cdot \beta_2^{-1} \cdot \ldots \cdot \alpha_g \cdot \beta_g \cdot \alpha_g^{-1} \cdot \beta_g^{-1}.$$



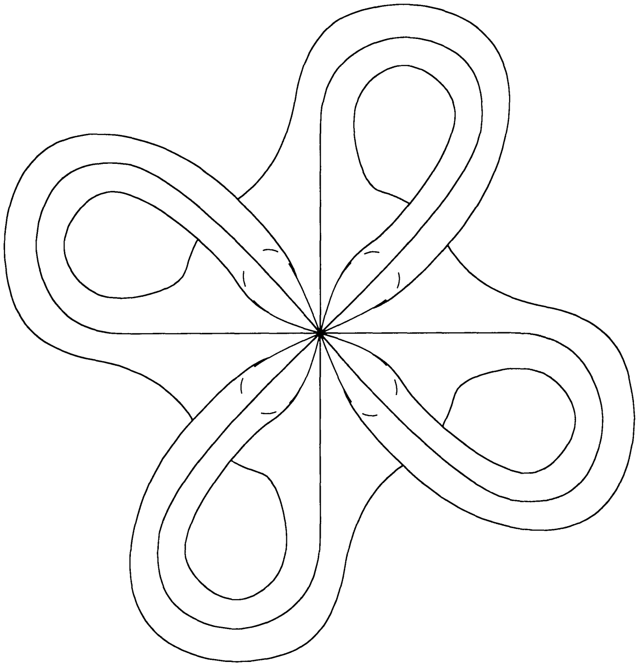
This is called a *normal form*. From this one can see directly that the identification space is a sphere with g handles. First make the identification of each β_i with β_i^{-1}:


α_3^{-1}
β_3
α_4
α_3
β_2
α_2^{-1}
α_4^{-1}
β_4
α_1
α_2
β_1
α_1^{-1}

or

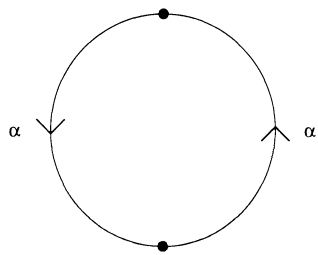

α_4^{-1}
α_3^{-1}
α_4
α_3
α_1
α_2
α_1^{-1}
α_2^{-1}

Then identify each α_i with α_i^{-1}:



Theorem 17.4. *If X is a (triangulable) compact orientable surface, then X is homeomorphic to a sphere with g handles for some nonnegative integer g.*

A similar procedure leads to a normal form for nonorientable compact surfaces. The projective plane can be realized by the code $\alpha \cdot \alpha$, identifying the opposite sides of a "two-sided" polygon



Problem 17.5. Show that a (triangulable) compact nonorientable sur-

face has a normal form given by a code $\alpha_1 \cdot \alpha_1 \cdot \alpha_2 \cdot \alpha_2 \cdot \ldots \cdot \alpha_h \cdot \alpha_h$, for some positive integer h.

For a discussion of triangulation of arbitrary compact surfaces, together with a more modern proof of Theorem 17.4, see Armstrong (1983).

17c. The Fundamental Group of a Surface

The normal form representation of a compact orientable surface X as a $4g$-sided polygon Π with sides identified according to the code (17.3) can be used to compute the fundamental group of X. Let x be the point in X that is the image of the vertices in Π, and let α_i and β_i be the loops in X that are the images of the corresponding sides of Π.

Let F_{2g} be the free group on $2g$ generators $a_1, b_1, \ldots, a_g, b_g$. There is a homomorphism of groups from F_{2g} to the fundamental group $\pi_1(X,x)$ that takes a_i to α_i and b_i to β_i for $1 \leq i \leq g$. We claim first that the element

$$c_g = a_1 \cdot b_1 \cdot a_1^{-1} \cdot b_1^{-1} \cdot a_2 \cdot b_2 \cdot a_2^{-1} \cdot b_2^{-1} \cdot \ldots \cdot a_g \cdot b_g \cdot a_g^{-1} \cdot b_g^{-1}$$

in F_{2g} maps to the identity element of $\pi_1(X,x)$ by this homomorphism. In fact, the product path

$$\alpha_1 \cdot \beta_1 \cdot \alpha_1^{-1} \cdot \beta_1^{-1} \cdot \alpha_2 \cdot \beta_2 \cdot \alpha_2^{-1} \cdot \beta_2^{-1} \cdot \ldots \cdot \alpha_g \cdot \beta_g \cdot \alpha_g^{-1} \cdot \beta_g^{-1}$$

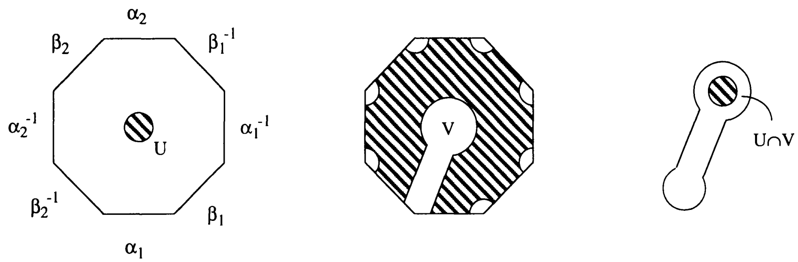
is homotopic to the constant path at x, since the corresponding path around the sides of Π is homotopic to a constant path in Π (since Π is convex), and composition with the continuous map from Π to X gives a homotopy in X.

Let N_g be the least normal subgroup of F_{2g} containing c_g, i.e., N_g is the subgroup generated by all elements of the form $u \cdot c_g \cdot u^{-1}$ for all u in F_{2g}. From what we have just seen, N_g is in the kernel of the homomorphism from F_{2g} to $\pi_1(X,x)$, so we have a homomorphism

$$F_{2g}/N_g \rightarrow \pi_1(X,x).$$

Proposition 17.6. *This homomorphism $F_{2g}/N_g \rightarrow \pi_1(X,x)$ is an isomorphism.*

Proof. We will apply the Van Kampen theorem. Let U be the image in X of the complement of a small disk in the middle of Π, and let V be the image in X of an open set that contains this disk as shown:



Let K be the image of the boundary of Π in X. This K is a graph that consists of $2g$ loops at x, so $\pi_1(K,x)$ is the free group on generators $a_1, b_1, \ldots, a_g, b_g$, i.e., $\pi_1(K,x) = F_{2g}$. By radial projection from the center of the disk, one sees that K is a deformation retract of U. We therefore have an isomorphism

$$\pi_1(U,x) \cong \pi_1(K,x),$$

so $\pi_1(U,x)$ is also the free group F_{2g} on the same generators. Now V is homeomorphic to a disk, so $\pi_1(V,x) = \{e_x\}$ is trivial, and $U \cap V$ has a circle for a deformation retract, so $\pi_1(U \cap V, x) = \mathbb{Z}$. The inclusion of $U \cap V$ in U takes a generator of $\pi_1(U \cap V, x)$ to the element $c_g = a_1 \cdot b_1 \cdot a_1^{-1} \cdot \ldots \cdot b_g^{-1}$. By Van Kampen's theorem, for any group G, to give a homomorphism from $\pi_1(X,x)$ to G is the same as giving a homomorphism from $\pi_1(U,x) = F_{2g}$ to G (and a homomorphism from $\pi_1(V,x) = \{e\}$ to G) in such a way that the composite

$$\mathbb{Z} = \pi_1(U \cap V, x) \to \pi_1(U,x) \to G$$

takes a generator of $\mathbb{Z}$ to the identity element of G. This means precisely that one has a homomorphism from F_{2g} to G such that c_g maps to the identity, or equivalently, that N_g maps to the identity. That is,

$$\mathrm{Hom}(\pi_1(X,x), G) = \mathrm{Hom}(F_{2g}/N_g, G).$$

This implies that the map from F_{2g}/N_g to $\pi_1(X,x)$ is an isomorphism, cf. Exercise 14.5. (See also Exercise 14.8, which applies to this situation directly.) □

Corollary 17.7. *The first homology group H_1X is a free abelian group of rank $2g$, with basis the image of the loops $\alpha_1, \beta_1, \ldots, \alpha_g, \beta_g$.*

Proof. Let A_{2g} be the free abelian group on $2g$ generators. The canonical map from F_{2g} to A_{2g} maps N_g to 0, since c_g is in the commutator subgroup. This determines a homomorphism from

$\pi_1(X,x) = F_{2g}/N_g$ to A_{2g}, and therefore a homomorphism

$$H_1(X) = \pi_1(X,x)/[\pi_1(X,x), \pi_1(X,x)] \to A_{2g}.$$

Since the images of $a_1, b_1, \ldots, a_g, b_g$ generate $H_1(X)$, and their images in A_{2g} are linearly independent, it follows that they are linearly independent in $H_1(X)$ and that this map is an isomorphism. □

In particular, since the rank of a free abelian group is an invariant of the group, this shows that the number g of Theorem 17.4 is independent of choices, and depends only on the topology of X. It is called the *genus* of the surface. Together with what we saw in Chapter 8, this shows that for any triangulation of X, the number v of vertices, e of edges, and f of faces satisfies the Euler equation

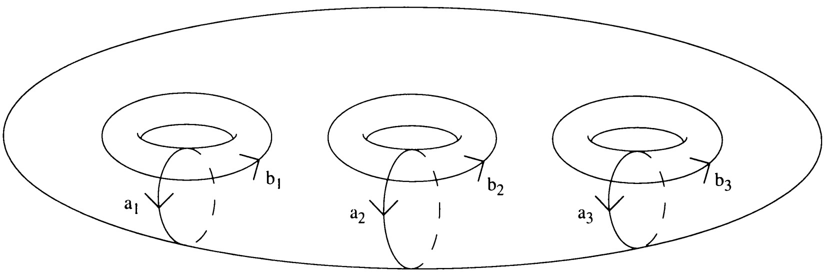
$$v - e + f = 2 - 2g.$$

Problem 17.8. For an nonorientable compact surface with normal form $\alpha_1 \cdot \alpha_1 \cdot \alpha_2 \cdot \alpha_2 \cdot \ldots \cdot \alpha_h \cdot \alpha_h$ as in Problem 17.5, show that the fundamental group is the quotient of the free group F_h on h generators $a_1, \ldots, a_h$ by the least normal subgroup that contains the element $a_1^2 \cdot a_2^2 \cdot \ldots \cdot a_h^2$. Show that the first homology group is isomorphic to $\mathbb{Z}^{\oplus(h-1)} \oplus (\mathbb{Z}/2\mathbb{Z})$. In particular, the number h is independent of all choices.

Exercise 17.9. Prove that for any triangulation of a compact nonorientable surface, with h as above, the number v of vertices, e of edges, and f of faces satisfies the Euler equation

$$v - e + f = 2 - h.$$

For homology there is no need to use paths with the same base points. It will be convenient to change the paths α_i and β_i to paths that we denote by a_i and b_i as shown:



Exercise 17.10. Show that the loops α_i and a_i define the same classes

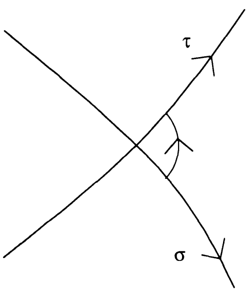
in $H_1(X)$, and similarly for β_i and b_i. In particular, the $2g$ classes determined by $a_1, b_1, \ldots, a_g, b_g$ form a free basis for $H_1(X)$.

The polygonal normal form of a surface can also be used to describe its universal covering space. When $g = 0$, of course, X is a sphere, so simply connected. When $g = 1$, X is a torus, which we have seen has the plane $\mathbb{R}^2$ as its universal covering. One can construct this universal covering from its representation as a rectangle with sides identified, by pasting the sides together but without identifying opposite sides. For $g \geq 2$, the universal covering can be realized in a similar way by gluing together copies of the polygon Π, but one cannot do this in a metrical way in the ordinary plane. However, in a hyperbolic plane this can be done, and one can see that the universal covering can be realized as a hyperbolic plane. The hyperbolic plane is homeomorphic to an open disk (or to $\mathbb{R}^2$), so from this one sees that for all $g \geq 2$ the universal covering is an open disk. For more on the hyperbolic plane, see Hilbert and Cohn-Vossen (1952).

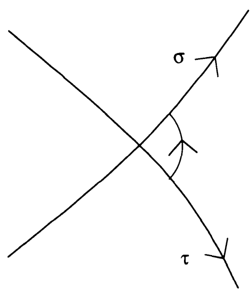
Exercise 17.11. Give another proof of the results of this section by computing the fundamental group of the complement of $2g$ disjoint disks in a two-sphere, and then showing what happens when one sews g handles (each homeomorphic to $S^1 \times [0, 1]$) onto pairs of the boundaries of these disks.

Problem 17.12. Compute the fundamental group and first homology group of the space obtained from a surface X of genus g by removing n disjoint disks or points.

There is another important operation on the first homology group H_1X of an oriented compact surface, an *intersection pairing*, that assigns a number $\langle\sigma, \tau\rangle$ to two classes σ and τ in H_1X. If the classes are represented by loops that meet each other transversally in a finite number of points, this number is the sum of the numbers ± 1 assigned to each point of intersection, with the number ± 1 assigned according to the following picture:



+1



-1

The number is $+1$ if the direction from σ to τ is counterclockwise, and -1 if it is clockwise. (Note that this notion depends on having an orientation for X.) What takes some proof is showing that this gives a well-defined number: that any two classes in H_1X have representatives that meet transversally (which is not difficult), and then that the number one gets is independent of choices (which is more difficult). Rather than doing this directly, we will construct this intersection pairing by another procedure in the next chapter, by relating it to the wedge product of 1-forms.

CHAPTER 18

Cohomology on Surfaces

18a. 1-Forms and Homology

On any $\mathscr{C}^\infty$ surface X, just as on an open set in the plane, we have a notion of $\mathscr{C}^\infty$ functions, 1-forms, and 2-forms ω, and we have linear maps d, that take a function f to a 1-form df, and a 1-form ω to a 2-form $d\omega$. (See Appendix D3 for definitions and basic properties.) Therefore we have a notion of a 1-form ω being *closed* ($d\omega = 0$) or *exact* ($\omega = df$ for some f). All exact forms are closed, so we can define the first De Rham cohomology group as before:

$$H^1X \;=\; \{\text{closed 1-forms on } X\}/\{\text{exact 1-forms on } X\}.$$

Just as in the case of open sets in the plane, if $\gamma\colon [a,b] \to X$ is a continuous path, and ω is a closed 1-form on X, we can define an integral $\int_\gamma \omega$. As before, if γ is differentiable, this can be done by calculus: $\int_\gamma \omega = \int_a^b \gamma^*\omega$. In general, as in the plane, we can subdivide the interval by $a = t_0 \le t_1 \le \ldots \le t_n = b$, so that $\gamma([t_{i-1}, t_i])$ is contained in an open set U_i on which $\omega = df_i$, and then

$$\int_\gamma \omega \;=\; \sum_{i=1}^{n} f_i(\gamma(t_i)) - f_i(\gamma(t_{i-1})).$$

The same argument as in §9b shows that this is independent of choices. As before, this extends from integrals over paths to integrals over 1-chains, and the same argument shows that the integral over a bound-

ary is zero, and that the integral of df over a cycle is zero. It follows that we again have a canonical map

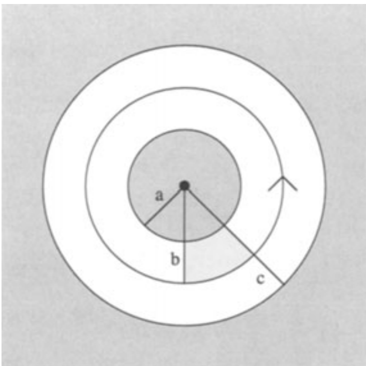
$$H^1X \rightarrow \mathrm{Hom}(H_1X, \mathbb{R})$$

that takes the class of ω to the homomorphism $[\gamma] \mapsto \int_\gamma \omega$.

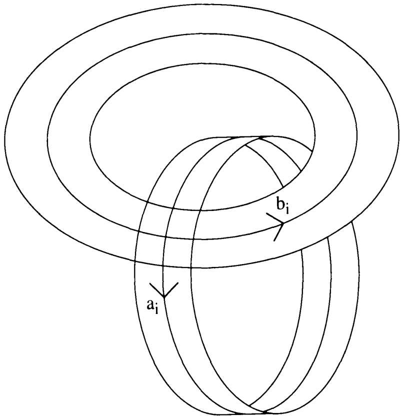
We want to calculate H^1X when X is a compact oriented surface. In particular, we want to show that this canonical map is an isomorphism. This could be done by the Mayer–Vietoris argument (which generalizes from open sets in the plane), but we will do it by explicitly constructing a basis, which will be useful later.

We take a model of X as in §17c, with a basis for H_1X given by the loops $a_1, b_1, \ldots, a_g, b_g$ as indicated in the picture there. For each loop a_i we will construct a closed 1-form α_i, and for each b_i a closed 1-form β_i, and we will show that the classes of these $2g$ 1-forms form a basis of H^1X. The forms α_i and β_i will depend on several choices, but we will see that their classes in H^1X depend only on the homology classes of the $2g$ loops $a_1, b_1, \ldots, a_g, b_g$.

For each of these loops we can find an open set V containing it that is diffeomorphic to an annulus $U = \{(x, y): a^2 < x^2 + y^2 < c^2\}$, with the loop corresponding to a counterclockwise circle around the middle of the annulus, i.e., to the path $t \mapsto (b\cos(2\pi t), b\sin(2\pi t))$, $0 \leq t \leq 1$, for some $a < b < c$:



We take these neighborhoods V to be disjoint, except for those containing the same a_i and b_i, which can be taken to intersect in a set diffeomorphic to an open rectangle.



Consider one such loop, either a_i or b_i, and fix such a diffeomorphism φ from a neighborhood V of the loop to such an annulus U. From Exercise B.14 we can find a $\mathscr{C}^\infty$ function ψ on the plane that is identically 1 outside the big circle and identically 0 inside the small circle. In fact, for any ε with $0 < \varepsilon < {}^1\!/_2(c-a)$, we can find ψ so that

$$\psi(x,y) = \begin{cases} 1 & \text{if } x^2+y^2 \geq (c-\varepsilon)^2, \\ 0 & \text{if } x^2+y^2 \leq (a+\varepsilon)^2. \end{cases}$$

The function $\psi \circ \varphi$ is a $\mathscr{C}^\infty$ function on V that is constant on a neighborhood of the boundary of V. Its differential $d(\psi \circ \varphi)$ is therefore a 1-form that is identically zero on a neighborhood of the boundary of V. We can therefore define a $\mathscr{C}^\infty$ 1-form on all of X by defining it to be $d(\psi \circ \varphi)$ on V, and identically 0 outside V. This is certainly a closed 1-form. In fact, its restriction to V is exact, and its restriction to an open set V' with $V \cup V' = X$ is zero. (Note that the form is *not* exact on X, however, as follows for example from the following lemma.) This is the 1-form we wanted to construct. We denote it by α_i if the given loop is a_i, and β_i if the given loop is b_i.

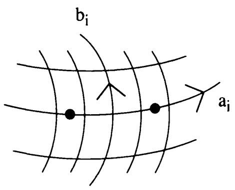
Lemma 18.1. *These* 1*-forms* α_i *and* β_i *have integrals*:

(i) $\displaystyle\int_{b_j} \beta_i = 0$ *for all* i *and* j, *and* $\displaystyle\int_{a_j} \beta_i = \begin{cases} 1 & \text{if } i=j, \\ 0 & \text{if } i \neq j; \end{cases}$

(ii) $\displaystyle\int_{a_j} \alpha_i = 0$ *for all* i *and* j, *and* $\displaystyle\int_{b_j} \alpha_i = \begin{cases} -1 & \text{if } i=j, \\ 0 & \text{if } i \neq j. \end{cases}$

Proof. This is a direct calculation. All of the assertions that integrals

vanish follow from the fact that they are integrals of 1-forms over a loop, and these forms are exact on an open set containing the loop. For the integral of β_i along a_i, cut a_i into two pieces, one inside the annulus as shown, the other outside.



For the piece inside the annulus, β_i is the differential of a function whose value at the initial point is 0 and at the end point is 1, so the integral of β_i along this piece is 1. For the second, β_i is identically zero. So the integral is $1 + 0 = 1$, as asserted. The argument is similar for the integral of α_i along b_i, but this time the function is 1 at the initial point and 0 at the final point of the part of the path in the annulus, so the integral is -1. □

Proposition 18.2. (1) *The canonical map* $H^1X \to Hom(H_1X, \mathbb{R})$ *is an isomorphism.*

(2) *The classes of the* 1*-forms* $\alpha_1, \beta_1, \ldots, \alpha_g, \beta_g$ *form a basis for* H^1X.

Proof. The same argument as in the plane shows that this canonical map is one-to-one. For if ω is a closed 1-form and $\int_\gamma \omega = 0$ for all closed paths γ, then one can define a function f on X by fixing a point P_0 and defining $f(P)$ to be the integral of ω along any path from P_0 to P. The assumptions make this a well-defined function, and the proof that $df = \omega$ is the same as for open sets in the plane (see Proposition 1.8); indeed, the assertion $df = \omega$ is a local assertion, so it can be verified on coordinate neighborhoods.

To see that the canonical map is subjective, a homomorphism h: $H_1X \to \mathbb{R}$ is determined by its values on a basis, so suppose $h([a_j]) = r_j$ and $h([b_j]) = s_j$ for some numbers r_j and s_j, $1 \le j \le g$. By the lemma, the closed 1-form $\omega = \Sigma_{i=1}^g (r_i \beta_i - s_i \alpha_i)$ has $\int_{a_j} \omega = r_j$ and $\int_{b_j} \omega = s_j$, which means that the canonical map takes ω to h. □

Exercise 18.3. Show that for closed 1-forms ω on the surface X one has a "module of periods" story as in Chapter 9: if the integrals of

ω along the basic loops are known, all integrals are determined. Show in fact that if $\int_{a_j}\omega = r_j$ and $\int_{b_j}\omega = s_j$, then for any closed 1-chain γ,

$$\int_\gamma \omega = \sum_{i=1}^{g} (m_i r_i + n_i s_i),$$

for some integers m_i and n_i.

Note in particular that the classes $[\alpha_i]$ and $[\beta_i]$ of α_i and β_i in H^1X are determined by their integrals, so are independent of the choices we made in defining them. They depend in fact only on the choice of the basis for H_1X determined by the a_i and b_i.

18b. Integrals of 2-Forms

Once an orientation of our surface is chosen, one can integrate $\mathscr{C}^\infty$ 2-forms. The integral of the 2-form ν on X will be a real number, denoted $\iint_X \nu$. More generally, if X is any oriented surface, and ν is a 2-form with compact support (i.e., which is identically zero outside a compact subset of X), one can define $\iint_X \nu$ by the following procedure. Choose an atlas of coordinate charts $\varphi_\alpha\colon U_\alpha \to X$, say with each U_α an open rectangle in the plane, with all charts compatible with the given orientation, and satisfying the condition that any point of X has a neighborhood meeting only finitely many $\varphi_\alpha(U_\alpha)$ (see Appendix A4 and Lemma 24.10).

To give a 2-form ν on X is the same as giving a 2-form $\nu_\alpha\, dx\, dy$ on U_α (which the same as prescribing a function ν_α on U_α), such that these 2-forms are compatible under changes of coordinates. First suppose ν is a 2-form that is zero except on a compact subset of one of the open sets $\varphi_\alpha(U_\alpha)$. Then one can define the integral of ν by

$$\iint_X \nu = \iint_{U_\alpha} \nu_\alpha\, dx\, dy,$$

where the integral on the right is the ordinary Riemann integral of a $\mathscr{C}^\infty$ (or continuous) function ν_α on a rectangle. In general, choose a partition of unity $\{\psi_\alpha\}$ so that the closure of the support of each ψ_α is contained in $\varphi_\alpha(U_\alpha)$. Define the integral of ν to be the sum of the integrals of the $\psi_\alpha \nu$, i.e.,

$$\iint_X \nu = \sum_\alpha \left(\iint_X \psi_\alpha \nu\right),$$

where the integral of each $\psi_\alpha \nu$ is defined by the preceding case, since it is zero except on $\varphi_\alpha(U_\alpha)$. Note by the compactness of the support of ν that only finitely many $\psi_\alpha \nu$ can be nonzero, so this sum is finite.

Exercise 18.4. (a) Verify that this definition is independent of the choice of partition of unity and the choice of coordinate charts. (b) Verify that the integral is linear: $\iint_X (r_1\nu_1 + r_2\nu_2) = \iint_X r_1\nu_1 + \iint_X r_2\nu_2$ for 2-forms ν_1 and ν_2 and real numbers r_1 and r_2. (c) Show that, if the other orientation is chosen, then the integral changes sign.

We need the following version of the Green–Stokes theorem for a surface:

Proposition 18.5. *If $\nu = d\omega$, where ω is a 1-form with compact support on an oriented surface X, then $\iint_X \nu = 0$.*

Proof. Choosing an atlas and partition of unity as in the definition, since $\omega = \Sigma_\alpha \psi_\alpha \omega$,

$$\iint_X \nu = \iint_X d\omega = \iint_X d\Big(\sum_\alpha \psi_\alpha\omega\Big) = \sum_\alpha \iint_X d(\psi_\alpha\omega).$$

Now $\psi_\alpha\omega$ is a 1-form with compact support on $\varphi_\alpha(U_\alpha)$, which corresponds to a 1-form μ_α with compact support on the open rectangle U_α. By definition, $\iint_X d(\psi_\alpha\omega) = \iint_{U_\alpha} d(\mu_\alpha)$, so it suffices to prove that this last integral is zero. But μ_α has compact support, so it vanishes on the boundary of the rectangle. By Green's theorem for a rectangle (Lemma 1.11),

$$\iint_{U_\alpha} d(\mu_\alpha) = \int_{\partial U_\alpha} \mu_\alpha = 0. \qquad \square$$

18c. Wedges and the Intersection Pairing

If X is a compact oriented surface, and ω and μ are 1-forms on X, we can define a real number (ω, μ) by the formula

$$(\omega, \mu) = \iint_X \omega \wedge \mu .$$

where $\omega \wedge \mu$ is the wedge product (Appendix D3). From basic properties of the wedge product (see Exercise D.7) we have:

(i) $(r_1\omega_1 + r_2\omega_2, \mu) = r_1(\omega_1, \mu) + r_2(\omega_2, \mu)$ for 1-forms ω_1, ω_2, and μ, and real numbers r_1 and r_2;
(ii) $(\omega, \mu) = -(\mu, \omega)$ for any 1-forms ω and μ; and
(iii) $(df, \mu) = 0$ for a function f and a closed 1-form μ.

Condition (i) says that (ω, μ) is linear in the first factor. Similarly, or using (ii), it is linear in the second, i.e., it is a *bilinear pairing*. Condition (ii) says that this pairing is *skew-symmetric*. Condition (iii), together with (ii), says that $(\omega, \mu) = 0$ if either ω or μ is exact. It follows that we can define a mapping

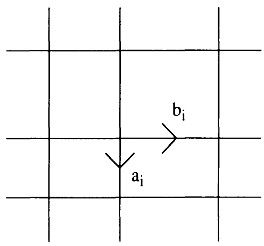
$$H^1X \times H^1X \to \mathbb{R}, \qquad [\omega] \times [\mu] \mapsto (\omega, \mu) = \iint_X \omega \wedge \mu .$$

The point is that, since the integral vanishes when either is exact, the result is independent of choice of a closed 1-form in an equivalence class. This is also a bilinear and skew-symmetric pairing.

Lemma 18.6. *Let α_i and β_i be the 1-forms constructed in* §18a. *Then $(\alpha_i, \alpha_j) = 0 = (\beta_i, \beta_j)$ for all i and j, and*

$$(\alpha_i, \beta_j) = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \neq j. \end{cases}$$

Proof. As in Lemma 18.1, this is a direct calculation. The pairings that are asserted to be 0 are evident, since in these cases the wedge products of the forms are zero. This uses the fact that a wedge product $\omega \wedge \mu$ vanishes where either ω or μ vanishes, and the fact that $\omega \wedge \omega = 0$ for any 1-form ω. It therefore suffices to verify that $(\alpha_i, \beta_i) = 1$, i.e., that $\iint_X \alpha_i \wedge \beta_i = 1$, for all i. The form $\alpha_i \wedge \beta_i$ vanishes off the region that is the intersection of the two annuli around the loops a_i and b_i, and this region is diffeomorphic to a rectangle. In a coordinate patch we have the picture



On this rectangle, $\alpha_i = df_i$ and $\beta_i = dg_i$, where f_i is a function that is 0 on the right side of the rectangle and 1 on the left side, and g_i is a

function that is 0 on the top of the rectangle and 1 on the bottom. If R is this rectangle,

$$\iint_X \alpha_i \wedge \beta_i = \iint_R df_i \wedge dg_i = \iint_R d(f_i \cdot dg_i) = \int_{\partial R} f_i \cdot dg_i,$$

the last step by Green's theorem on the rectangle R. Now since dg_i vanishes on the bottom and top edges of the rectangle, $f_i \cdot dg_i$ is zero on all sides of the rectangle except for the left vertical edge γ_4, where it is equal to dg_i. So

$$\int_{\partial R} f_i \cdot dg_i = -\int_{\gamma_4} dg_i = -(g_i(\gamma_4(1)) - g_i(\gamma_4(0))) = -(0-1) = 1,$$

which finishes the proof. □

Exercise 18.7. For any closed 1-form ω on X, show that

$$(\alpha_j, \omega) = \int_{a_j} \omega \quad \text{and} \quad (\beta_j, \omega) = \int_{b_j} \omega.$$

Exercise 18.8. Show that, for any closed 1-forms μ and ν,

$$(\mu, \nu) = \sum_{j=1}^{g} \left(\int_{a_j} \mu \int_{b_j} \nu - \int_{a_j} \nu \int_{b_j} \mu \right).$$

Proposition 18.9. *The pairing $H^1X \times H^1X \to \mathbb{R}$ is a perfect pairing, i.e., for any linear map φ: $H^1X \to \mathbb{R}$, there is a unique ω in H^1X such that $\varphi(\mu) = (\omega, \mu)$ for all μ in H^1X.*

Taking $\varphi = 0$, this says in particular that if $\omega \in H^1X$ and $(\omega, \mu) = 0$ for all μ in H^1X, then $\omega = 0$.

Proof. In this proof we identify closed 1-forms with the classes they define in H^1X. If $\omega = \Sigma_{i=1}^{g}(r_i\alpha_i + s_i\beta_i)$ in H^1X, then by Lemma 18.6 we have

$$(\omega, \beta_j) = \sum_{i=1}^{g} (r_i(\alpha_i, \beta_j) + s_i(\beta_i, \beta_j)) = r_j,$$

$$(\omega, \alpha_j) = -(\alpha_j, \omega) = -\sum_{i=1}^{g} (r_i(\alpha_j, \alpha_i) + s_i(\alpha_j, \beta_i)) = -s_j,$$

The homomorphism determined by ω therefore takes α_j to $-s_j$ and β_j to r_j. If this homomorphism is zero, all s_j and r_j must be zero, which means that ω is zero in H^1X. Conversely, any homomorphism φ is

determined by the values it takes on a basis of H^1X, and if we define r_j to be $\varphi(\beta_j)$ and $s_j = -\varphi(\alpha_j)$, the same equations show that $\omega = \Sigma_{i=1}^{g}(r_i\alpha_i + s_i\beta_i)$ is a class in H^1X with $(\omega, \mu) = \varphi(\mu)$ for all μ. □

Corollary 18.10. *For any γ in H_1X, there is a unique class ω_γ in H^1X such that $\int_\gamma \mu = \iint_X \omega_\gamma \wedge \mu$ for all μ in H^1X.*

Proof. Given γ, the map $\mu \mapsto \int_\gamma \mu$ is a homomorphism from H^1X to $\mathbb{R}$, so the proposition gives a unique class ω_γ with the required property. □

Exercise 18.11. (a) For $\gamma = a_i$, the corresponding class ω_{a_i} is represented by the form α_i; and for $\gamma = b_i$, the corresponding class ω_{b_i} is represented by the form β_i. This shows again that the classes of the forms α_i and β_i depend only on the choice of $a_1, \ldots, b_g$. (b) Show that the map from H_1X to H^1X taking γ to ω_γ is a homomorphism.

Problem 18.12. Suppose γ is represented by a map from the circle S^1 to X that extends to a diffeomorphism from an annulus containing S^1 to an open subset of X. Use this annulus as in §17a to construct a differential form ω that vanishes outside the image of the annulus, and, on the annulus, is the differential of a function that increases from 0 to 1 as one moves from the inside to the outside of the annulus. Show that ω represents the class ω_γ of the corollary.

We can use the corollary to define an *intersection number* $\langle\sigma, \tau\rangle$ for any σ and τ in H_1X, as we indicated at the end of the last chapter. Namely, let ω_σ and ω_τ be the cohomology classes given by the corollary, and define $\langle\sigma, \tau\rangle$ to be $(\omega_\sigma, \omega_\tau)$, i.e.,

$$\langle\sigma, \tau\rangle = \iint_X \omega_\sigma \wedge \omega_\tau = \int_\sigma \omega_\tau = -\iint_X \omega_\tau \wedge \omega_\sigma = -\int_\tau \omega_\sigma .$$

Proposition 18.13. (1) *This pairing is a bilinear skew-symmetric pairing on H_1X, i.e.,*

$$\begin{aligned} \langle m_1\sigma_1 + m_2\sigma_2, \tau\rangle &= m_1\langle\sigma_1, \tau\rangle + m_2\langle\sigma_2, \tau\rangle, \\ \langle\sigma, m_1\tau_1 + m_2\tau_2\rangle &= m_1\langle\sigma, \tau_1\rangle + m_2\langle\sigma, \tau_2\rangle, \\ \langle\tau, \sigma\rangle &= -\langle\sigma, \tau\rangle, \end{aligned}$$

for all σ, τ, σ_1, σ_2, τ_1, τ_2 in H_1X and all integers m_1 and m_2.

(2) *With the basis* a_i *and* b_i *for* H_1X *as before,* $\langle a_i, a_j\rangle = \langle b_i, b_j\rangle = 0$ *for all* i *and* j*, and*

$$\langle a_i, b_j\rangle = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \neq j. \end{cases}$$

(3) *The number* $\langle \sigma, \tau\rangle$ *is always an integer.*

(4) *For any* σ *in* H_1X*, the map* $\tau \to \langle \sigma, \tau\rangle$ *is a homomorphism from* H_1X *to* $\mathbb{Z}$*, and every homomorphism from* H_1X *to* $\mathbb{Z}$ *arises in this way from a unique* σ *in* H_1X.

Proof. Part (1) follows from the corresponding assertions for the pairing (,) on cohomology classes. Part (2) follows from Lemma 18.6 and Exercise 18.11: for example, $\langle a_i, b_j\rangle = (\alpha_i, \beta_j)$, which is 1 if $i = j$ and 0 otherwise. Since the a_i and b_i form a basis for H_1X, (3) follows from (1) and (2). The proof of (4) is the same as before: the element $\sigma = \Sigma r_i a_i + s_i b_i$ has $\langle \sigma, a_j\rangle = -s_j$ and $\langle \sigma, b_j\rangle = r_j$, with this time r_i and s_i integers, and a homomorphism is determined by arbitrarily specifying the integers to which each element of a basis maps. □

Problem 18.14. Suppose σ and τ are represented by maps from S^1 to X that extend to diffeomorphisms of annuli with open sets in X, so that the images cross transversally at a finite number of points. Show how to assign numbers $+1$ or -1 to each intersection point (see the end of §17), in such a way that $\langle \sigma, \tau\rangle$ is the sum of these numbers.

It should be pointed out that, as we have constructed it here, the intersection pairing uses a differentiable structure on the surface. However, as the preceding problem indicates; it really depends only on the topology.

Problem 18.15. Prove this assertion.

18d. De Rham Theory on Surfaces

We have concentrated on the first De Rham group H^1X. As in the plane, on any surface one has the 0th group H^0X, which is the space of locally constant functions on X. If X is connected, such functions are constant, and H^0X is just the space $\mathbb{R}$ of constant functions.

We can define the *second De Rham group* H^2X of a differentiable surface X by

$$H^2X = \{\text{2-forms on } X\}/\{\text{exact 2-forms on } X\}.$$

If X is a compact oriented surface, it follows from Proposition 18.5 that integration defines a canonical map

$$H^2X \to \mathbb{R}, \qquad [\nu] \mapsto \iint_X \nu.$$

It is easy to see that this map is surjective, by constructing a form ν that vanishes outside one coordinate neighborhood $\varphi_\alpha(U_\alpha)$, and in U_α has an expression $v_\alpha\, dx\, dy$, with v_α a nonnegative function with given integral. We claim that, in fact, this canonical map is an isomorphism. This is equivalent to the

Claim 18.16. *If* $\iint_X \nu = 0$, *then* ν *must be exact.*

Equivalently, *the dimension of* H^2X *is* (at most) *one*. With this, we will know all the De Rham cohomology groups of a compact surface of genus g:

$$H^0X = \mathbb{R}, \qquad H^1X \cong \mathbb{R}^{\oplus 2g}, \qquad H^2X \cong \mathbb{R}.$$

We will show how to extend the Mayer–Vietoris story to include the second cohomology group, which will in particular prove this claim. We will also see that if X is not compact, or if X is not orientable, then $H^2X = 0$: every closed 2-form is exact.

Exercise 18.17. Show directly from the definition that, if U is an open rectangle, every 2-form is exact. Conclude that $H^2U = 0$ if U is diffeomorphic to an open rectangle.

Exercise 18.18. Suppose a surface X is a disjoint union of a finite or infinite number of open sets U_i. Show that specifying a k-form on X is the same as specifying a k-form on each U_i, and specifying a class in H^kX is the same as specifying a class in H^kU_i for each i. In other words, H^kX is the direct product of the groups H^kU_i.

We want to compare the cohomology groups for different open sets. Note first that if U_1 is a subset of U_2, any differential function or form on U_2 determines by restriction a differential function or form on U_1. This restriction commutes with the boundary maps d (which amounts to the obvious fact that partial derivatives of a function re-

stricted from a larger open set to a subset are the restrictions of the partial derivatives). This means that restriction takes closed forms to closed forms, and exact forms to exact forms, and hence determines linear maps, also called *restriction* maps:

$$H^k(U_2) \to H^k(U_1), \qquad k = 0,1,2.$$

Exercise 18.19. If $U_1 \subset U_2 \subset U_3$, show that the restriction map from $H^k(U_3)$ to $H^k(U_1)$ is the composite of the restriction map from $H^k(U_3)$ to $H^k(U_2)$ followed by the restriction map from $H^k(U_2)$ to $H^k(U_1)$.

Exercise 18.20. If U and V are open sets in a surface X, and ω is a k-form on $U \cap V$ (with $k = 1$ or 2), show that there are k-forms ω_1 on U and ω_2 on V so that $\omega = \omega_1 - \omega_2$ on $U \cap V$.

Given open sets U and V in a surface X, the contructions of Chapter 10 extend without change, giving a coboundary map

$$\delta\colon H^0(U \cap V) \to H^1(U \cup V).$$

The Mayer–Vietoris properties MV(i)–MV(v) of §10d continue to hold for these spaces and maps. Similarly, we construct a coboundary map

$$\delta\colon H^1(U \cap V) \to H^2(U \cup V).$$

The construction of a map from {closed 1-forms on $U \cap V$} to $H^2(U \cup V)$ proceeds exactly as before: Given a closed 1-form ω on $U \cap V$, use Exercise 18.20 to write ω as $\omega_1 - \omega_2$, with ω_1 and ω_2 1-forms on U and V, respectively. The 2-forms $d\omega_1$ and $d\omega_2$ agree on $U \cap V$, so there is a unique 2-form μ on $U \cup V$ that agrees with $d\omega_1$ on U and with $d\omega_2$ on V. Define $\delta(\omega)$ to be the class of this 2-form μ: $\delta(\omega) = [\mu]$. Exactly as before, one checks that this is independent of the choice of ω_1 and ω_2, and is a linear map. To see that it defines a homomorphism on the quotient space

$$H^1(U \cap V) = \{\text{closed 1-forms on } U \cap V\}/\{\text{exact 1-forms on } U \cap V\},$$

we must show that the map just defined vanishes on the exact 1-forms. If $\omega = df$, write $f = f_1 - f_2$, with f_1 and f_2 functions on U and V, respectively. Then in our construction of the coboundary of ω we may take $\omega_1 = df_1$ and $\omega_2 = df_2$, from which it follows that $d\omega_1$ and $d\omega_2$ are both zero, so $\mu = 0$ and $\delta([\omega]) = 0$, as required.

We claim next that properties (i)–(v) continue:

MV(vi). *Given ω in $H^1(U \cap V)$, $\delta(\omega) = 0$ if and only if $\omega = \alpha - \beta$ for some α in H^1U and β in H^1V.*

MV(vii). *Given μ in $H^2(U \cup V)$, μ restricts to zero in H^2U and H^2V if and only if $\mu = \delta(\omega)$ for some ω in $H^1(U \cap V)$.*

MV(viii). *Given α in H^2U and β in H^2V, α and β have the same restriction to $H^2(U \cap V)$ if and only if α and β are the restrictions of some element in $H^2(U \cup V)$.*

The proofs are exactly the same as before, and are left as exercises. In addition, we have, as an immediate consequence of Exercise 18.20:

MV(ix). *Any class μ in $H^2(U \cap V)$ can be written as the difference of a class α in H^2U and a class β in H^2V.*

Putting this together, using the fancy language as before, we have the full

Theorem 18.21 (Mayer–Vietoris Theorem). *For any open sets U and V in a surface, there is an exact sequence*

$$\begin{aligned} 0 &\longrightarrow H^0(U \cup V) \xrightarrow{+} H^0U \oplus H^0V \xrightarrow{-} H^0(U \cap V) \\ &\xrightarrow{\delta} H^1(U \cup V) \xrightarrow{+} H^1U \oplus H^1V \xrightarrow{-} H^1(U \cap V) \\ &\xrightarrow{\delta} H^2(U \cup V) \xrightarrow{+} H^2U \oplus H^2V \xrightarrow{-} H^2(U \cap V) \longrightarrow 0. \end{aligned}$$

Problem 18.22. Use Mayer–Vietoris to show that $H^2X = 0$ for any open subset of the plane.

Proposition 18.23. *If X is a surface diffeomorphic to a sphere with g handles, then $\dim H^2X = 1$.*

Proof. We can write X as the union of the same two open sets U and V that we used in the proof of Proposition 17.6. Since V is diffeomorphic to an open disk, $H^1V = 0 = H^2V$, see Exercise 18.17. We leave it as an exercise to show similarly that $H^2U = 0$. We know that $H^1(U \cap V) \cong \mathbb{R}$. The proof of Proposition 17.6 shows that the map from H_1U to H_1X induced by inclusion is an isomorphism. It follows from this and Proposition 18.2 that the restriction map from H^1X to H^1U is an isomorphism. The relevant part of the Mayer–Vietoris sequence is therefore

$$H^1X \xrightarrow{\cong} H^1U \oplus 0 \rightarrow H^1(U \cap V) \rightarrow H^2X \rightarrow 0 \oplus 0.$$

It follows that the map from $H^1(U \cap V)$ to H^2X is an isomorphism. Since $H^1(U \cap V) \cong \mathbb{R}$, this shows that H^2X is one dimensional. □

Exercise 18.24. Give another proof of the proposition by writing X as a union of a sphere with $2g$ disjoint disks removed, and the union of g handles, each diffeomorphic to a cylinder $S^1 \times (0, 1)$.

Exercise 18.25. If X is a nonorientable compact surface, use Mayer–Vietoris to show that $H^2X = 0$. If X is written in normal form as in Problem 17.5, show that the dimension of H^1X is $h - 1$. Show that for all the compact surfaces, if triangulated with v vertices, e edges, and f faces,

$$v - e + f = \dim(H^0X) - \dim(H^1X) + \dim(H^2X).$$

Problem 18.26. (a) If $f\colon Y \to X$ is a differentiable map of surfaces, show how to define pull-backs $f^*\omega$ of forms from X to Y, determining maps $f^*\colon H^kX \to H^kY$ for $k = 0, 1, 2$. (b) If f is an n-sheeted covering map, show how to define push-forwards $f_*\omega$ of forms from Y to X, defining $f_*\colon H^kY \to H^kX$. (c) With f as in (b), show that the composite $f_* \circ f^*\colon H^kX \to H^kX$ is multiplication by n. (d) If X is a compact non-orientable surface, and $f\colon \tilde{X} \to X$ is its orientation covering, show that $\iint_{\tilde{X}} f^*\omega = 0$ for any 2-form ω on X. Conclude again that $H^2X = 0$.

We will see in Chapter 24 that there are corresponding homology groups H_2X, with a corresponding Mayer–Vietoris sequence, and isomorphisms $H^2X \cong \mathrm{Hom}(H_2X, \mathbb{R})$. We will also see in Chapter 24 that H^2X vanishes for any noncompact surface.

PART X

RIEMANN SURFACES

Many of the ideas about covering spaces, homology, and cohomology, can be used in the study of Riemann surfaces. A Riemann surface is a differentiable surface with a complex analytic structure. Compact Riemann surfaces arise by taking a finite-sheeted covering of the complement of a finite set in the two-sphere S^2, and then filling in appropriately over the branch points. We prove the Riemann–Hurwitz formula that computes the genus (number of handles) of a surface arising this way.

If $F(z, w)$ is an irreducible polynomial in two complex variables, it defines w as an algebraic function of z, and one is interested in the behavior of integrals of the form $\int dz/w$, or more generally $\int R(z, w)\, dz$, where R is any rational function. Associated with F there is a complex plane curve $F(z, w) = 0$; as a subset of $\mathbb{C}^2$ it is, except for a possible finite number of singularities, a Riemann surface, and the above process compactifies this surface. One of the key discoveries of the nineteenth century was how the topology of this surface, i.e., the genus g, controls much of the analysis related to the algebraic function w and its integrals. The integrals between two points are defined up to "periods," which are integrals along classes in the first homology group of the Riemann surface.

In Chapter 21 we use what we have learned about the topology and differential forms on a surface to prove the celebrated Riemann–Roch theorem, which shows how that topology (genus) of a surface controls the kinds of meromorphic functions and forms one can find on the surface. We also prove the Abel–Jacobi theorem, which similarly re-

lates the possible integrals on the surface to its topology. We prove these theorems only for Riemann surfaces that are known to arise from a plane curve, i.e., from an irreducible polynomial $F(Z, W)$. (It is true that all compact Riemann surfaces arise this way, but we do not prove this here.) In this case there is a short proof of Weil that uses the algebraic notion of adeles, together with the few facts from analysis and topology that we have available.

CHAPTER 19

Riemann Surfaces

19a. Riemann Surfaces and Analytic Mappings

A *Riemann surface* X is a connected surface with a special collection of coordinate charts $\varphi_\alpha: U_\alpha \to X$. As before, U_α is a subset of $\mathbb{R}^2$, but now we identify $\mathbb{R}^2$ with the complex numbers $\mathbb{C}$. The requirement to be a Riemann surface is that the change of coordinate mappings $\varphi_{\beta\alpha}$ from $U_{\alpha\beta} \subset U_\alpha$ to $U_{\beta\alpha} \subset U_\beta$ are not just $\mathscr{C}^\infty$, but they must also be *analytic*, or *holomorphic*. Recall (see §9d) that an analytic function f on an open set in $\mathbb{C}$ is a complex-valued function that is locally expandable in a power series, i.e., at each point z_0 in the open set, there is a power series $\Sigma_{n=0}^{\infty} a_n(z - z_0)^n$ that converges to $f(z)$ for all z in some neighborhood of z_0. As before, another atlas of charts is compatible with a given one (and defines the same Riemann surface) if the changes of coordinates from charts in one to charts in the other are all analytic.

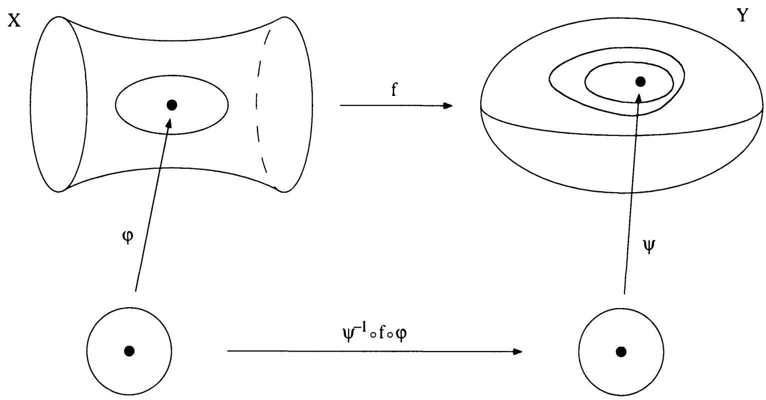
Any Riemann surface has a natural orientation. This follows from the fact that, when the analytic changes of coordinates are written out in terms of their real and imaginary parts, the Jacobian of the result has positive determinant. We will consider only connected Riemann surfaces.

Exercise 19.1. Show that in fact, if $w = f(z)$, with $w = u + iv$ and $z = x + iy$, the determinant of the Jacobian of the map from (x, y) to (u, v) is the square of the absolute value of the complex derivative $f'(z)$.

We have seen several examples of Riemann surfaces. Of course, $\mathbb{C}$ itself is one, with the identity coordinate chart $\mathbb{C} \to \mathbb{C}$, and any open subset of $\mathbb{C}$ (or any Riemann surface) is also a Riemann surface. The sphere S^2 is a compact Riemann surface, with the charts given by spherical projection from the north and south poles. Indeed, we saw (Exercise 7.14) that the change of coordinates mapping in this case is given by the map $z \mapsto 1/z$ from $\mathbb{C} \setminus \{0\}$ to $\mathbb{C} \setminus \{0\}$. (When we speak of S^2, it will always be regarded as a Riemann surface with these coordinate charts.) The torus $\mathbb{R}^2/\mathbb{Z}^2 = \mathbb{C}/(\mathbb{Z} + \mathbb{Z}i)$ is a compact Riemann surface, using the projections from small open sets in $\mathbb{C}$ to the quotient for coordinate charts.

Any compact oriented surface can be given a structure of a Riemann surface (as we shall see later), but, except for the sphere S^2, there are infinitely many nonequivalent Riemann surfaces with the same underlying surface. For example, if $\Lambda \subset \mathbb{C}$ is any lattice, i.e., a subgroup generated by two elements that give a basis for $\mathbb{C}$ as a real vector space, then $\mathbb{C}/\Lambda$ is likewise a compact Riemann surface. These are all homeomorphic (and diffeomorphic) to each other, but they are generally not isomorphic Riemann surfaces (see Exercise 19.17). This is part of the general subject of "moduli of Riemann surfaces"—the study of the set of all Riemann surfaces of a given genus—that we can only hint at in this book.

If X is a Riemann surface, it makes sense to say that a function $f: X \to \mathbb{C}$ is *analytic* (or *holomorphic*): f is analytic if for each coordinate chart $\varphi_\alpha: U_\alpha \to X$, the composite $f \circ \varphi_\alpha$ is an analytic function on the open set U_α in $\mathbb{C}$. More generally, if X and Y are Riemann surfaces, a mapping $f: X \to Y$ is *analytic* at a point P in X if there are charts $\varphi: U \to X$ and $\psi: V \to Y$ mapping to neighborhoods of P and $f(P)$, respectively, so that $f(\varphi(U)) \subset \psi(V)$, and the composite $\psi^{-1} \circ f \circ \varphi$ is an analytic function from U to V:



This condition is independent of the choice of coordinates φ and ψ. In fact, we can choose both U and V to be disks centered at the origins, with these origins mapped to P and $f(P)$. In this case the composite $h = \psi^{-1} \circ f \circ \varphi$ has the form $h(z) = \Sigma_{n=1}^{\infty} a_n z^n$, for some converging power series. As we will soon see, the order of vanishing of h at the origin, i.e., the smallest integer e such that $a_e \neq 0$, is independent of the choice of coordinates. This integer is called the *ramification index* of f at P, and we will denote it by $e_f(P)$, or just $e(P)$ when one function f is being considered. (If f is constant, of course, h is identically 0, and $e = \infty$, but we will not be interested in constant maps.) The point P is called a *ramification point* for f if $e_f(P) > 1$.

We claim next that we can change the coordinate chart φ so that the composite $\psi^{-1} \circ f \circ \varphi$ is the function $z \mapsto z^e$. To see this, write $h(z) = z^e \cdot g(z)$, where $g(z)$ is an analytic function at the origin and $g(0) \neq 0$. ($g(z) = \Sigma_{n=0}^{\infty} a_{n+e} z^n$.) Such an analytic function g, possibly in a smaller disk around the origin, can be written as the eth power of an analytic function $k(z)$; for example, $g(z)/a_e$ maps 0 to 1, so maps a neighborhood of 0 to the right half-plane, so one can compose it with a branch of the log function

$$\log(z) \;=\; -\sum_{n=1}^{\infty} \frac{1}{n}(1-z)^n \qquad \text{for} \quad |z-1| < 1\,,$$

and then set

$$k(z) \;=\; \alpha \cdot \exp\!\left(\frac{1}{e}\log(g(z)/a_e)\right),$$

where α is any eth root of a_e and $\exp(z) = \Sigma_{n=0}^{\infty}(1/n!)z^n$.

Exercise 19.2. Verify that $k(z)^e = g(z)$ in some disk containing the origin. Show that there are exactly e choices for such k (up to shrinking the disks they are defined on), obtained by the e choices for the eth root α of a_e.

Now we have written $h(z) = (z \cdot k(z))^e$, where k is an analytic function with $k(0) \neq 0$. The mapping $z \mapsto z \cdot k(z)$ is an analytic isomorphism in some neighborhood of the origin, since its derivative does not vanish at the origin. Therefore, we can define a new coordinate chart $\tilde{\varphi}$ (from some small disk U' to X) so that $\tilde{\varphi}(z) = \varphi(z \cdot k(z))$ for all small z. It follows that, with this new coordinate chart $\tilde{\varphi}$, the composite $\psi \circ f \circ \tilde{\varphi}$ is just the map $z \mapsto z^e$, as required.

The mapping $z \mapsto z^e$ is familiar: it maps 0 to 0, and outside the

origin it is an e-sheeted covering map. This shows that there are neighborhoods U of P and V of $f(P)$ so that f maps U to V, and the mapping $U \setminus \{P\} \to V \setminus \{f(P)\}$ is an e-sheeted covering. This shows in particular that the number e is independent of choices, and in fact depends only on the topology of the map f near P.

Note another important consequence of this local structure: a nonconstant analytic mapping is always an *open* mapping, i.e., the image of any open set is open. In particular, if X is compact and Y is not compact, the only analytic mappings from X to Y are constants. When $Y = \mathbb{C}$ this openness is a strong form of the maximum principal: one cannot have a point P_0 such that $|f(P_0)| \geq |f(P)|$ for P in a neighborhood of P_0.

Proposition 19.3. *Let $f: X \to Y$ be a nonconstant analytic map between compact Riemann surfaces.*

(1) *There are a finite number of ramification points. Let $R \subset X$ be the set of ramification points, and set $S = f(R) \subset Y$.*

(2) *The map from $X \setminus f^{-1}(S)$ to $Y \setminus S$ determined by f is an n-sheeted covering map for some finite number n. This integer n is called the degree of f.*

(3) *For any point Q in Y, $\Sigma_{P \in f^{-1}(Q)} e_f(P) = n$.*

Proof. (1) follows from the fact that for any point P, there is a neighborhood of P that contains no other ramification point, using the compactness of X to cover it by a finite number of such neighborhoods. It follows similarly that $f^{-1}(Q)$ is finite for all points Q in Y (consider a possible limit point of the set $f^{-1}(Q)$).

To prove (2), let $Q \notin S$, and $f^{-1}(Q) = \{P_1, \ldots, P_n\}$. There are neighborhoods U_i of P_i and V_i of Q such that f maps U_i homeomorphically onto V_i. Shrinking the U_i if necessary, we may assume they are disjoint, and that V_i contains no point of S. For any connected neighborhood V of Q contained in $V_1 \cap \ldots \cap V_n$, let $U_i' = U_i \cap f^{-1}(V)$; note that f maps each U_i' homeomorphically onto V. We claim that for V sufficiently small, $f^{-1}(V)$ is the disjoint union of the sets U_i', from which it follows that V is evenly covered, so f is a covering in a neighborhood of Q. To prove this claim, suppose on the contrary that there is a sequence of neighborhoods N_i of Q whose intersection is $\{Q\}$, such that there is a point P_i' in $f^{-1}(N_i)$ with P_i' not in $U_1 \cup \ldots \cup U_n$. By the compactness of X, a subsequence of these P_i' must have a limit point P' in X. By the continuity of f, $f(P') = Q$, so $P' = P_j$ for some j. But this contradicts the assumption that the points

P_i' are not in U_j for all i and j. (In fact, having proved (2), it follows that the whole neighborhood $V = \cap V_i$ is evenly covered.)

For (3), again let $f^{-1}(Q) = \{P_1, \ldots, P_m\}$, and find neighborhoods U_i of P_i and V_i of Q such that f maps U_i onto V_i, so that in local coordinates it is the map $z \mapsto z^{e(P_i)}$, so it is $e_f(P_i)$ to 1 except at the point P_i. Again if V is a neighborhood of Q contained in the intersection of the V_i, but not containing any other point of S, then there are $\Sigma e_f(P_i)$ points over a point Q' in V except for the point Q. By (2), this sum must be the number n of sheets of the covering. □

A *meromorphic function* f on a Riemann surface X is the same as an analytic function $f: X \to S^2$ from X to the Riemann sphere. Equivalently, f is an analytic function from $X \setminus S$ to $\mathbb{C}$, with S a discrete subset of X, and for each point P in X, there is a coordinate chart $\varphi: U \to X$ taking 0 to P, so that on $U \setminus \{P\}$, $f \circ \varphi(z) = z^k \cdot g(z)$, for some integer k and some analytic function g that is nonvanishing in U. The integer k is called the *order* of f at P, and is denoted $\mathrm{ord}_P(f)$. It is independent of the choice of the local coordinate. If k is positive, one says that f has a *zero* of order k, or *vanishes to order* k, and if k is negative, we say that f has a *pole of order* $-k$.

Exercise 19.4. Show that f is meromorphic at P if and only if there is a coordinate chart as above and an integer k so that the function $z^{-k} \cdot (f \circ \varphi)(z)$ is bounded as z approaches 0.

In terms of the mapping $f: X \to S^2$, the order of f at P is positive if $f(P) = 0$, negative if $f(P) = \infty$, and zero otherwise. If $f(P) = 0$, the order of f at P is just the ramification index $e_f(P)$ of f at P. Similarly, if $f(P) = \infty$, the order of f at P is minus the ramification index at P, since $1/z$ is a local parameter for the Riemann sphere at ∞. If X is compact, we know that the sum of the ramification indices over any point in S^2 is the degree n of f. Assertion (3) of the proposition says that f takes on all values the same number of times, counting multiplicity correctly. In particular, for the values 0 and ∞, this gives:

Corollary 19.5. *For any nonconstant meromorphic function f on a compact Riemann surface X,*

$$\sum_{P \in X} \mathrm{ord}_P(f) = 0.$$

Exercise 19.6. Let $p_1(z)$ and $p_2(z)$ be polynomials in z of degrees d_1 and d_2, with no common factors. (a) Show that $f(z) = p_1(z)/p_2(z)$ de-

termines an analytic mapping from $S^2 = \mathbb{C} \cap \{\infty\}$ to itself. (b) Show that the degree of this mapping is the maximum of the integers d_1 and d_2. (c) Show that the ramification index of this mapping at ∞ is $|d_1 - d_2|$. (d) Show that every analytic mapping from S^2 to S^2 has this form.

19b. Branched Coverings

If Y is a Riemann surface, and $p: X \to Y$ is a covering mapping, then there is a unique structure of a Riemann surface on X so that p is an analytic mapping. In fact, one can choose charts $\varphi_\alpha: U_\alpha \to Y$ for Y so that each $\varphi_\alpha(U_\alpha)$ is evenly covered by p; each component $V_{\alpha,i}$ of $p^{-1}(\varphi_\alpha(U_\alpha))$ maps homeomorphically to $\varphi_\alpha(U_\alpha)$ by p, and the composite

$$U_\alpha \to \varphi_\alpha(U_\alpha) \to V_{\alpha,i}$$

of φ_α and the inverse of p, gives a coordinate chart on X. It is straightforward to verify that the changes of coordinates for these charts are analytic, so define a Riemann surface structure on X.

Exercise 19.7. If Y is a Riemann surface, and G is a group acting evenly on Y by analytic isomorphisms, show that Y/G can be given the structure of a Riemann surface so that $Y \to Y/G$ is analytic.

An important case of this is the fact that the universal covering $\tilde{X}$ of a Riemann surface is a Riemann surface, and the fundamental group $\pi_1(X, x)$ acts as a group of analytic isomorphisms of $\tilde{X}$. When $X = S^2$, $\tilde{X} = X$; when $X = \mathbb{C}/\Lambda$ for a lattice Λ, $\tilde{X} = \mathbb{C}$. It is a fact of complex analysis (the *uniformization theorem*) that in all other cases, the universal covering is isomorphic to the upper half plane H (or an open disk). The automorphisms of H all have the form $z \mapsto (az+b)/(cz+d)$ where $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ is a real matrix of determinant 1. This means that X can be realized as the quotient of H by a subgroup of $\mathrm{SL}_2(\mathbb{R})/\{\pm I\}$ acting evenly on H. We will not have more to say about this situation, which is a fundamental area of mathematics in its own right.

In the last section we saw that any (nonconstant) analytic map $X \to Y$ of compact Riemann surfaces determines a finite-sheeted covering $X \setminus f^{-1}(S) \to Y \setminus S$, for some finite set S in Y. Our goal in this section is to reverse this process. Given a Riemann surface Y, a finite subset S of Y, and a finite-sheeted topological covering $p: X^\circ \to Y \setminus S$, by what we just proved, X° gets a structure of a Riemann surface so that this

mapping is analytic. We want to "fill in" the missing points over the points of S, embedding X° in a Riemann surface X, so that we have an analytic mapping f from X to Y compatible with the given covering:

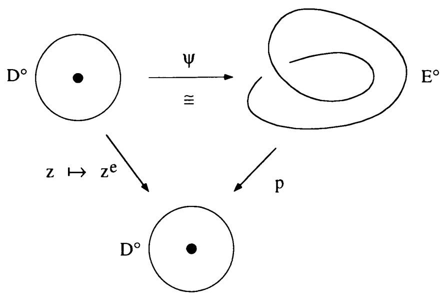
$$\begin{array}{ccc} X^\circ & \subset & X \\ {\scriptstyle p}\downarrow & & \downarrow{\scriptstyle f} \\ Y\setminus S & \subset & Y. \end{array}$$

If Y is compact, we want X to be compact. In general, we want the mapping f to be *proper*: for any compact set K of Y, the inverse image $f^{-1}(K)$ should be a compact set in X. The problem is local on Y: we need to fill in the covering over each point of S. We will look first at the "local model" of a Riemann surface—an open disk.

Let $D = \{z: |z| < 1\}$, $D^\circ = D \setminus \{0\}$. We know all about the coverings of D°. In fact, its fundamental group is $\mathbb{Z}$ (since it contains a circle as a deformation retract), so connected finite-sheeted coverings correspond to subgroups of finite index in $\mathbb{Z}$, and the only such subgroups of $\mathbb{Z}$ are the groups $e\mathbb{Z}$ for e a positive integer. The e-sheeted covering corresponding to this subgroup is

$$p_e: D^\circ \to D^\circ, \qquad z \mapsto z^e.$$

This means that if $p: E^\circ \to D^\circ$ is any e-sheeted connected covering, there is a homeomorphism $\psi: D^\circ \to E^\circ$ so that $p \circ \psi = p_e$.

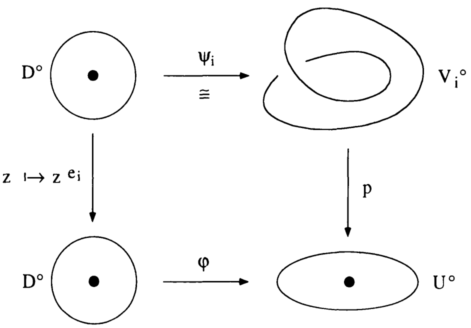


This homeomorphism ψ is not uniquely determined, depending, as we know, on a choice of where a base point is mapped. In particular, there are exactly e such homeomorphisms, corresponding to the e points of E° over a given base point. The other choices of ψ have the form $z \mapsto \psi(\zeta_k \cdot z)$, where $\zeta_k = \exp(2\pi k i/e)$ is one of the eth roots of unity, $1 \le k \le e-1$.

It is now clear how to fill in the covering $p: E^\circ \to D^\circ$. Define E to be the union of E° with one other point, and put the structure of a

Riemann surface on E so that the extension of ψ from D° to D, mapping 0 to the added point, is an isomorphism of D with E. Note that this does not depend on the choice of ψ, since the map $z \mapsto \zeta_k \cdot z$ is an analytic isomorphism.

Now return to the given covering $p: X^\circ \to Y \setminus S$. Given Q in S, we can find a coordinate chart $\varphi: D \to \varphi(D) = U \subset Y$, with D the open unit disk in the plane, and $\varphi(0) = Q$, such that U does not contain any other point of S but Q. Let $U^\circ = U \setminus \{Q\}$. The covering p restricts to a covering of U°, so $p^{-1}(U^\circ)$ is a disjoint union of connected open sets $V_1^\circ, \ldots, V_m^\circ$, with each $V_i^\circ \to U^\circ$ a connected covering, say with e_i sheets. By what we just saw, one can find homeomorphisms $\psi_i: D^\circ \to V_i^\circ$ such that the diagram



commutes, i.e., $p(\psi_i(z)) = \varphi(z^{e_i})$, with ψ_i unique up to first multiplying z by an e_ith root of unity. We can therefore add one point to each V_i°, getting spaces V_i so that each ψ_i extends to a homeomorphism from D to V_i. Taking these extensions as charts, the disjoint union of the V_i becomes a Riemann surface. The map from V_i° to U° extends to an analytic map from V_i to U that has ramification index e_i at the added point.

If this is done at each point of S, one gets a space X that is the union of X° with a finite number of points. These local charts give X the structure of a Riemann surface (noting that the added charts are compatible with the given charts on X°), and the covering p is extended to an analytic mapping $f: X \to Y$.

Exercise 19.8. Verify that this defines a Riemann surface X, and the map f is analytic and proper.

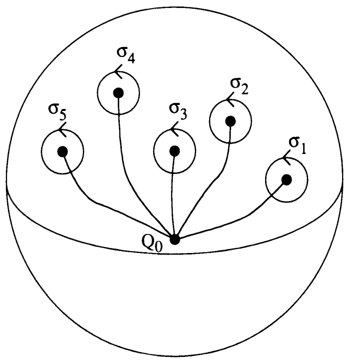
These results are summarized in the following proposition:

Proposition 19.9. *Let Y be a Riemann surface, S a finite subset of Y, and $p: X^\circ \to Y \setminus S$ a finite-sheeted covering map, with X° connected. Then there is an embedding of X° as an open subset of a Riemann surface X that is a union of X° and a finite set, so that p extends to a proper analytic mapping from X to Y.*

Exercise 19.10. Show that X, with its Riemann surface structure, is unique up to canonical isomorphism: if $X^\circ \subset X' \to Y$ is another, there is a unique isomorphism of Riemann surfaces from X to X' compatible with the inclusions of X° and mappings to Y.

Exercise 19.11. Extend the proposition to the case where S is an infinite but discrete subset of Y.

One reason for the importance of this proposition is that, since we know the fundamental group of $Y \setminus S$ (see Problem 17.12 for the general case), we know all possible finite-sheeted coverings. The main case of concern to us will be $Y = S^2$, with S a set of r points. By choosing disjoint arcs from a base point Q_0 to these points, we can number the points $S = \{Q_1, \ldots, Q_r\}$ so that the arcs to them occur in this order, going counterclockwise around Q_0. Then one can construct loops $\sigma_1, \ldots, \sigma_r$ that go from Q_0 along an arc to a point near Q_i, make a counterclockwise circle around Q_i, and go back along the arc to Q_0.



Exercise 19.12. Show that $\pi_1(S^2 \setminus \{Q_1, \ldots, Q_r\}, Q_0)$ is the free group F_r on the generators $\sigma_1, \ldots, \sigma_r$, modulo the least normal subgroup containing $\sigma_1 \cdot \ldots \cdot \sigma_r$. This is isomorphic to a free group on $r-1$ generators $\sigma_1, \ldots, \sigma_{r-1}$.

Remark 19.13. In §16d we described a correspondence between n-sheeted coverings of a space and actions of its fundamental group on a finite set T with n elements. To give an action of this fundamental group on a set T is the same as giving r permutations $s_1, \ldots, s_r$ of T, with the requirement that $s_1 \cdot \ldots \cdot s_r$ is the identity permutation. So any such data determine a branched covering of the sphere. The classical way to do this was to use the fact that the complement of the r arcs drawn is simply connected, so a covering over this is a disjoint union of r copies of it. Labeling these by elements of T, the permutations are used to describe how to glue the sheets together across the arcs, to get the Riemann surface X. (Note, however, that the arcs are only a tool used to give nice generators for the fundamental group, and are not necessary for describing the covering.)

Exercise 19.14. (a) Carry out this construction, and show that it agrees with that described before. (b) Show that the covering corresponding to this choice of permutations is connected if and only if they generate a transitive subgroup of the automorphisms of T. (Recall that a group of permutations is *transitive* if, for any elements t_1 and t_2 of T, there is a permutation in the group that takes t_1 to t_2.) (c) Let X be the corresponding Riemann surface, with $f\colon X \to S^2$ the analytic mapping. For each i, write T as a disjoint union of sets $T_{i,j}$ so that s_i permutes the elements of each $T_{i,j}$ cyclically. Show that there is one point of X over Q_i for each set $T_{i,j}$, and the cardinality of $T_{i,j}$ is the ramification index of f at this point.

19c. The Riemann–Hurwitz Formula

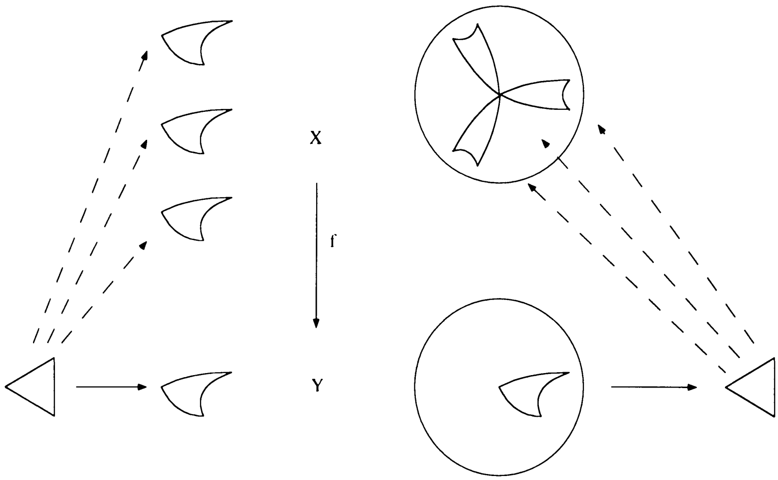
Given an analytic mapping $f\colon X \to Y$ between compact Riemann surfaces, our next aim is to describe the topology of X in terms of the topology of Y and the local behavior around the branch points.

Theorem 19.15 (Riemann–Hurwitz). *Let $f\colon X \to Y$ be an analytic map of degree n between compact Riemann surfaces. If Y can be triangulated, so can X, and the genus g_X of X and the genus g_Y of Y are related by the formula*

$$2g_X - 2 \;=\; n(2g_Y - 2) + \sum_{P \in X} (e_f(P) - 1)\,.$$

Proof. By refining the triangulation of Y if necessary, for example, using the barycentric subdivision described in §8b, we can assume the triangulation has the property that each point Q of S is a vertex in the triangulation. We will construct from this an explicit triangulation of X. It is simplest to describe the vertices and the "open" edges (homeomorphic to $(0, 1)$) and the open faces (homeomorphic to the interior of a plane triangle) of this triangulation of X: these are exactly the components of the inverse images by f of the corresponding vertices, open edges, and open faces of the triangulation of Y. The point is that each of the open edges and faces on Y is in the locus where the map is a covering, and, since edges and faces are simply connected, the restriction of the covering to each of them is trivial, which means that there are n subsets of X that map homeomorphically onto each of them, and these are the specified open edges and faces.

For each closed edge and face on Y the triangulation specifies a homeomorphism with a closed interval or a closed triangle. Composing with the projections, we have compatible homeomorphisms of the open edges and faces on X with the interiors of these intervals or triangles. To see that we have a triangulation, we must verify that these homeomorphisms extend to the closed intervals and triangles. This is clear when the image edges or triangles in Y do not have vertices in S, since in that case the coverings are trivial over the entire closed triangle. The same argument shows that the homeomorphisms extend (uniquely, by continuity) to the closures of the intervals and triangles, except possibly in an arbitrarily small neighborhood of a vertex mapping to a point in $f^{-1}(S)$. To see that it extends continuously to such points is a local question, so we can assume we are in the situation $z \mapsto z^e$ of the mapping from the disk to itself. The corresponding vertex is sent to the center of the disk, and the continuity of the map at the vertex follows from the fact that $|z| \to 0$ if and only if $|z^e| \to 0$. The following picture shows the two types of behavior:



Now suppose the triangulation of Y has v vertices, e edges, and f faces. By the construction of the triangulation of X, it has $n \cdot f$ faces and $n \cdot e$ edges. But the number of vertices is not $n \cdot v$, since over each point Q of S there may be fewer than n points. In fact, by the equation $\Sigma e_f(P) = n$, the sum taken over the points P in $f^{-1}(Q)$, the number of points over Q is $n - \Sigma(e_f(P) - 1)$. It follows that the number of vertices in the triangulation of X is $n \cdot v - \Sigma(e_f(P) - 1)$, with the sum over all P in $f^{-1}(S)$, or in X, since $e_f(P) = 1$ if P is not in $f^{-1}(S)$. The Euler characteristic of X is therefore

$$n \cdot v - \sum_{P \in X} (e_f(P) - 1) - n \cdot e + n \cdot f \; = \; n(v - e + f) - \sum_{P \in X} (e_f(P) - 1).$$

Replacing the left side by $2 - 2g_X$, and $v - e + f$ by $2 - 2g_Y$, the formula of the proposition results. □

Exercise 19.16. Give an alternative proof of the theorem by triangulating Y so that each point of S lies inside a face, and no face contains two points of S.

For example, when $Y = S^2$, we have Riemann's formula

$$\sum_{P \in X} (e_f(P) - 1) \; = \; 2g_X + 2n - 2.$$

If X is also the sphere, this gives

$$\sum_{P \in X} (e_f(P) - 1) \; = \; 2n - 2.$$

The Riemann–Hurwitz formula can be useful to limit the possibilities for mappings between Riemann surfaces. Consider for example a nonconstant analytic mapping $f\colon \mathbb{C}/\Lambda \to \mathbb{C}/\Lambda'$, where Λ and Λ' are two lattices in $\mathbb{C}$. Since both have genus 1, it follows from the Riemann–Hurwitz formula that $e_f(P) = 1$ for all P, i.e., f must be unramified. Now f lifts to an analytic mapping $\tilde{f}\colon \mathbb{C} \to \mathbb{C}$ so that the diagram

$$\begin{array}{ccc} \mathbb{C} & \xrightarrow{\tilde{f}} & \mathbb{C} \\ {\scriptstyle p}\downarrow & & \downarrow{\scriptstyle p'} \\ \mathbb{C}/\Lambda & \xrightarrow{f} & \mathbb{C}/\Lambda', \end{array}$$

commutes. In fact, since $\mathbb{C}$ is simply connected, the composite $f \circ p$ lifts through the covering p' by Proposition 13.5, to produce a continuous map $\tilde{f}$, and $\tilde{f}$ is automatically analytic since the projections from $\mathbb{C}$ to the quotient spaces are local isomorphisms. Similarly, since f is unramified, $\tilde{f}$ is also unramified. From the topology one can see also that $\tilde{f}$ extends continuously to a map from $S^2 = \mathbb{C} \cap \{\infty\}$ to itself, taking ∞ to ∞. From Exercise 19.4 this extension, also denoted $\tilde{f}$, is an analytic mapping from S^2 to S^2. If the degree of $\tilde{f}$ is n, the sum of the $e_f(P) - 1$ can be at most $n - 1$ (since ramification can take place only over ∞), and this contradicts the Riemann–Hurwitz formula unless $n = 1$. By Exercise 19.6, we deduce that $\tilde{f}(z) = \lambda z + \mu$ for some complex numbers λ and μ, with $\lambda \neq 0$. In particular, $\lambda \cdot \Lambda + \mu \subset \Lambda'$. Conversely, any such λ and μ determine an analytic mapping f. This puts strong restrictions on the relations between the lattices, and on the possible maps f.

Exercise 19.17. (a) Show that for any lattice Λ there is a nonzero complex number a so that $a \cdot \Lambda$ is generated by 1 and τ, where τ is a number in the upper half plane. So every $\mathbb{C}/\Lambda$ is isomorphic to one of the form $\mathbb{C}/(\mathbb{Z} + \mathbb{Z} \cdot \tau)$. (b) Show that for two numbers τ and τ' in the upper half plane, $\mathbb{C}/(\mathbb{Z} + \mathbb{Z} \cdot \tau)$ is isomorphic to $\mathbb{C}/(\mathbb{Z} + \mathbb{Z} \cdot \tau')$ if and only if $\tau' = (a\tau + b)/(c\tau + d)$ for some integers a, b, c, d with $ad - bc = 1$. (c) Show that the only analytic maps from the Riemann surface $\mathbb{C}/(\mathbb{Z} + \mathbb{Z} \cdot \tau)$ to itself that take the image of 0 to itself are given by multiplication by an integer n, unless τ satisfies a quadratic polynomial with integer coefficients.

It follows from the preceding exercise that the set of compact Riemann surfaces that are isomorphic to some $\mathbb{C}/\Lambda$ are in one-to-one correspondence with the space $\mathrm{H}/\mathrm{SL}_2(\mathbb{Z})$ of orbits of the discrete group

$SL_2(\mathbb{Z})$ on the upper half plane H. In particular, not all such Riemann surfaces are isomorphic as Riemann surfaces, although they are all diffeomorphic to $S^1 \times S^1$. We will see later that every compact Riemann surface of genus 1 has the form $\mathbb{C}/\Lambda$, so $H/SL_2(\mathbb{Z})$ is the moduli space of compact Riemann surfaces of genus 1.

Exercise 19.18. Let $f: X \to Y$ be a nonconstant analytic map between compact Riemann surfaces. Show that $g_X \geq g_Y$, and $g_X > g_Y$ unless $g_Y = 0$ or $g_Y = 1$ or f is an isomorphism.

CHAPTER 20

Riemann Surfaces and Algebraic Curves

20a. The Riemann Surface of an Algebraic Curve

If $F(Z, W)$ is a polynomial in two variables, with complex coefficients, that is not simply a constant, its zero set

$$C = \{(z, w) \in \mathbb{C}^2 \colon F(z, w) = 0\}$$

is called a "complex affine plane curve." Identifying $\mathbb{C}^2$ with $\mathbb{R}^4$, C is defined by two real equations: the vanishing of the real and imaginary parts of $F(z, w)$. We may therefore expect C to be a surface, and this expectation is generally true, except that, just as in the case of real curves, C may have singularities. We will use the construction of the preceding section to "remove" these singularities, and also add some points over them and "at infinity," to get a compact Riemann surface. In fact, if F is not irreducible, the surface we get will be the disjoint union of the surfaces we get from the irreducible factors of F, so we assume for now that F is an irreducible polynomial, i.e., it has no nontrivial factors but constants. Write

$$F(Z, W) = a_0(Z)W^n + a_1(Z)W^{n-1} + \ . \ . \ . \ + a_{n-1}(Z)W + a_n(Z),$$

with $a_i(Z)$ a polynomial in Z alone, and $a_0(Z) \neq 0$. We may also assume that n is positive, for otherwise $F = bZ + c$, and C is isomorphic to $\mathbb{C}$, given by the projection to the second factor. We will need a little piece of algebra. Let $F_w = \partial F/\partial W$, which is a polynomial of

degree $n-1$ in W:

$$F_w = \frac{\partial F}{\partial W} = n \cdot a_0(Z)W^{n-1} + (n-1) \cdot a_1(Z)W^{n-2} + \ldots + a_{n-1}(Z).$$

Lemma 20.1. *There are polynomials* $B(Z, W)$, $C(Z, W)$, *and* $d(Z)$, *with* $d(Z) \neq 0$, *so that*

$$B(Z, W) \cdot F(Z, W) + C(Z, W) \cdot F_w(Z, W) = d(Z).$$

Proof. We use the lemma of Gauss (Appendix C3): if F is irreducible in $\mathbb{C}[Z, W]$, then F is also irreducible in $\mathbb{C}(Z)[W]$, where $\mathbb{C}(Z)$ is the field of rational functions in Z. The equation of Lemma 20.1 can be found (and computed) from the Euclidean algorithm in the ring $\mathbb{C}(Z)[W]$, as follows. Divide F by F_w to get, after clearing denominators, an equation of polynomials

$$b_0 \cdot F = Q_1 \cdot F_w + R_1,$$

where $b_0 \in \mathbb{C}[Z]$, and R_1 is a polynomial of degree less than $n-1$ in W. If the degree of R_1 in W is positive, divide F_w by R_1, getting an equation

$$b_1 \cdot F_w = Q_2 \cdot R_1 + R_2;$$

continuing, find equations $b_i \cdot R_{i-1} = Q_{i+1} \cdot R_i + R_{i+1}$, until for some k, R_{k+1} has degree 0 in W. Note that $R_{k+1} \neq 0$, since otherwise R_k would divide R_{k-1}, then $R_{k-2}, \ldots$, and finally F_w and F, contradicting the fact that F is irreducible.

Set $d = R_{k+1}$. To find an equation of the form required amounts to showing that d is in the ideal in $\mathbb{C}[Z, W]$ generated by F and F_w; this ideal contains R_1 by the first equation, then R_2 by the second, and so on, until finally it contains $R_{k+1} = d$. □

If one takes an equation as in the lemma so that B, C, and d are not all divisible by any nonconstant polynomial in Z, then d will be unique up to multiplication by a constant; this d is called the *discriminant* of F with respect to W. For our purposes any d will do.

Exercise 20.2. For $F = W^2 + b(Z)W + c(Z)$, show that $d(Z) = b(Z)^2 - 4c(Z)$. For $F = W^3 + b(Z)W + c(Z)$, show that $d(Z) = 4b(Z)^3 + 27c(Z)^2$.

Let p be the first projection from C to $\mathbb{C}$: $p(z, w) = z$. We will show that, if a finite number of points are removed, p becomes a covering map.

Lemma 20.3. *There is a finite subset S of $\mathbb{C}$ such that the projection from $C \setminus p^{-1}(S)$ to $\mathbb{C} \setminus S$ is a finite covering with n sheets.*

Proof. From Lemma 20.1 we draw the conclusion that if $z \in \mathbb{C}$ and $d(z) \neq 0$, then there is no w with $F(z, w) = 0$ and $F_w(z, w) = 0$. This means that the equation $F(z, W) = 0$ has no multiple roots. If in addition $a_0(z) \neq 0$, then this equation has n distinct roots. We take S to be the set where $a_0(z) \cdot d(z) \neq 0$.

Suppose $z_0 \notin S$, and let $w_1, \ldots, w_n$ be the roots of the equation $F(z_0, W) = 0$. We want to find analytic functions $g_1, \ldots, g_n$ defined in a neighborhood of z_0 so that $g_i(z_0) = w_i$ and $F(z, g_i(z)) \equiv 0$. One way to do this is to apply the implicit function theorem, either for a subset of $\mathbb{R}^4$ defined by two real equations, or the complex analogue for a subset of $\mathbb{C}^2$ defined by one equation. Another is to use the Argument Principle. For this, take small disjoint closed disks around these n points, and let γ_i be a counterclockwise path around the boundary of the disk around w_i. For z near z_0 and w on a circle γ_i, $F(z, w) \neq 0$ by continuity. For z near z_0,

$$\frac{1}{2\pi i}\int_{\gamma_i} \frac{F_w(z, w)}{F(z, w)}\, dw = 1,$$

since the integral is an integer that is 1 when $z = z_0$, and it varies continuously with z. This means that there is exactly one root of the equation $F(z, W) = 0$ inside γ_i. In fact, from Problem 9.34, this root is given by the formula

$$g_i(z) = \frac{1}{2\pi i}\int_{\gamma_i} w\frac{F_w(z, w)}{F(z, w)}\, dw.$$

In particular, take U to be a disk around z_0 where all these functions g_i are defined, and set $V_i = \{(z, g_i(z)) \colon z \in U\}$. Since these points give all possible roots of $F(z, W) = 0$ over z in U, we see that $p^{-1}(U)$ is the union of open sets V_i, and p maps each V_i homeomorphically onto V, with an inverse given by $z \mapsto (z, g_i(z))$. □

Now regard $\mathbb{C} \subset S^2 = \mathbb{C} \cup \{\infty\}$ as usual, and enlarge S by including the point at infinity in it. By the lemma, we have a covering map $X^\circ \to S^2 \setminus S$, with $X^\circ = \{(z, w) \colon F(z, w) = 0 \text{ and } z \notin S\}$. This covering map gives X° the structure of a Riemann surface, except that we must prove it is connected:

Lemma 20.4. *If $F(Z, W)$ is irreducible, then X° is connected.*

Proof. Let Y° be a connected component of X°. If X° is not connected,

then the covering $X^\circ \to S^2 \setminus S$ restricts to a covering $Y^\circ \to S^2 \setminus S$ with $m < n$ sheets (see Exercise 11.11). For each z not in S, let $e_1(z)$, $e_2(z), \ldots, e_m(z)$ be the elementary symmetric functions in the m values of w on the points in $p^{-1}(z) \cap Y^\circ$. That is, $e_1(z)$ is the sum of these m values, $e_2(z)$ is the sum of all products of pairs of these values, and so on until $e_m(z)$ is the product of all m values. These functions e_i are clearly analytic on $S^2 \setminus S$. They are in fact meromorphic on S^2, as follows from the fact that, for a in S, after multiplying by some $(z-a)^k$, the roots $w_i(z)$ approach 0 (see Exercise 19.4). Consider the polynomial

$$G = W^m - e_1(Z)W^{m-1} + e_2(Z)W^{m-2} - \ldots + (-1)^m e_m(Z)$$

in $\mathbb{C}(Z)[W]$. For any z not in S,

$$G(z, W) = \prod_{P \in p^{-1}(z) \cap Y^\circ} (W - w(P)),$$

$$F(z, W) = a_0(z) \prod_{P \in p^{-1}(z)} (W - w(P)).$$

It follows that G divides F in $\mathbb{C}(Z)[W]$, since the remainder obtained by dividing F by G would be a polynomial whose coefficients are rational functions of Z vanishing on an infinite set. This contradicts the irreducibility of F, and completes the proof. □

By Proposition 19.9 this covering can be extended to a compact Riemann surface X, together with an analytic map $f\colon X \to S^2$. This Riemann surface is called *the Riemann surface of the algebraic curve C* or of the polynomial F. It is easily seen to be independent of choice of S. In fact, the following exercise shows more.

Exercise 20.5. A point $P = (z_0, w_0)$ is called a *nonsingular* point of C if either: (i) $F_w(z_0, w_0) \neq 0$; or (ii) $F_z(z_0, w_0) \neq 0$; or both. (a) Show that the canonical map from X° to C extends to an isomorphism of a neighborhood in X with C in a neighborhood of P. (b) Show that the ramification index of $z\colon X \to S^2$ at the corresponding point is 1 in case (i), and the order of vanishing at w_0 of the function $w \mapsto F(z_0, w)$ in case (ii).

It is a theorem in analysis that every compact Riemann surface is the Riemann surface of an algebraic curve. The problem for proving this is to produce the meromorphic functions "z" and "w" on X. It is not at all obvious that there are any nonconstant meromorphic functions. For those that arise as branched coverings of S^2 (which is the

same as producing one nonconstant meromorphic function z), it is not obvious how to produce another not in $\mathbb{C}(z)$; once one produces a meromorphic function w that takes n distinct values on the n points in X over some given point of S^2, however, it is not hard to see that w satisfies some irreducible equation $F(z, w) \equiv 0$, and that X is the Riemann surface of this polynomial.

Let us work out the example with $F(Z, W) = W^3 + Z^3W + Z$. By Exercise 20.2, $d(Z) = 4Z^9 + 27Z^2$, so the possible branch points are where $z = 0$, $z = \infty$, and the solutions of $z^7 = -27/4$. The points of C over the finite points are nonsingular. Over 0 there is one point $(0, 0)$, and the ramification index of $z: X \rightarrow S^2$ is 3. Over each of the seventh roots of $-27/4$ there are two points, so one of each must have ramification index 2 and the other must be unramified. The point at infinity can be analyzed by making the substitution $Z' = 1/Z$, but one can see from the Riemann–Hurwitz formula that the sum of the numbers $e_z(P) - 1$ for the points P over ∞ must be odd, so there must be two points over ∞, with ramification indices 2 and 1. The genus g_X of the Riemann surface is given by

$$2g_X - 2 \;=\; -2n + \Sigma(e_z(P) - 1) \;=\; -6 + 2 + 7 \cdot 1 + 1\,,$$

so the genus is 3.

Exercise 20.6. For each of the following polynomials, compute a set S as in Lemma 20.3, compute the ramification indices of the points over S in the Riemann surface, and compute the genus of the Riemann surface: (i) $W^2 - \Pi_{i=1}^m(Z - a_i)$, $a_1, \ldots, a_m$ distinct complex numbers; (ii) $4W^3 - 3Z^2W + Z^3 - 2Z$; (iii) $W^3 - Z^6 + 1$; (iv) $W^3 - 3W^2 + Z^6$; and (v) $W^m + Z^m + 1$.

Exercise 20.7. (a) If a compact Riemann surface X has a meromorphic function with only one pole of order 1, show that X is isomorphic to S^2. (b) If X has a meromorphic function with two poles of order 1, or one pole of order 2, show that it gives a two-sheeted covering $X \rightarrow S^2$. (c) Show that, for a given set of $2g + 2$ points in S^2, there is, up to isomorphism, exactly one two-sheeted covering that branches at these points.

20b. Meromorphic Functions on a Riemann Surface

The meromorphic functions on a Riemann surface X form a field, which we denote by $M(X)$ or just M. If X is the Riemann surface of

the polynomial $F(Z, W)$, then the functions z and w (which come from the two projections of $C \subset \mathbb{C}^2$ to the axes) are seen as in Lemma 20.4 to be meromorphic functions on X, as are any rational functions of z and w. Such rational functions form a subfield of M, which we can denote by $\mathbb{C}(z, w)$.

Proposition 20.8. *Every meromorphic function on X is a rational function of z and w*:

$$\begin{aligned} M &= M(X) = \mathbb{C}(z, w) \\ &= \mathbb{C}(z) + \mathbb{C}(z) \cdot w + \mathbb{C}(z) \cdot w^2 + \ldots + \mathbb{C}(z) \cdot w^{n-1}. \end{aligned}$$

Proof. Note first that any element in $\mathbb{C}(z)[w]$ can be written as a polynomial of degree at most $n-1$ is w, as seen by dividing by F. So it suffices to show that any meromorphic function h on X is in $\mathbb{C}(z)[w]$.

From Lemma 20.1 we have an equation

$$\begin{aligned} d(z) \cdot h &= B(z, w) \cdot F(z, w) \cdot h + C(z, w) \cdot F_w(z, w) \cdot h \\ &= C(z, w) \cdot F_w(z, w) \cdot h, \end{aligned}$$

so it suffices to show that $F_w(z, w) \cdot h$ is in $\mathbb{C}(z)[w]$. For z not in the branch set S, let $P_1, \ldots, P_n$ be the points of X over z. Then

$$F(z, W) = a_0(z) \cdot \prod_{j=1}^{n} (W - w(P_j)),$$

$$F_w(z, W) = a_0(z) \cdot \sum_{i=1}^{n} \prod_{j \neq i} (W - w(P_j)),$$

so, for $1 \leq k \leq n$,

$$h(P_k) \cdot F_w(z, w(P_k)) = a_0(z) \cdot h(P_k) \cdot \prod_{j \neq k} (w(P_k) - w(P_j)).$$

Now consider the expression

$$a_0(z) \cdot \sum_{i=1}^{n} h(P_i) \cdot \prod_{j \neq i} (T - w(P_j)).$$

On the one hand, as in Lemma 20.4, this can be written in the form $\sum_{m=0}^{n-1} b_m(z) T^m$ with each b_m meromorphic on S^2, so $b_m \in \mathbb{C}(z)$. On the other hand, the preceding calculation shows that

$$\begin{aligned} h(P_k) \cdot F_w(z, w(P_k)) &= a_0(z) \cdot h(P_k) \cdot \prod_{j \neq k} (w(P_k) - w(P_j)) \\ &= \sum_{m=0}^{n-1} b_m(z) w(P_k)^m. \end{aligned}$$

This means that $h \cdot F_w(z,w)$ and $\Sigma b_m(z)w^m$ agree on the complement of a finite set, which implies that they are equal. □

Exercise 20.9. If $F_z(z_0, w_0) \neq 0$, show that the ramification index of $z\colon X \to S^2$ at the point corresponding to (z_0, w_0) is one more than the order of the meromorphic function $F_w(z,w)$ at (z_0, w_0).

With $Y = S^2$ and $f\colon X \to Y$ the mapping given by z, and with $p\colon X^\circ \to Y^\circ = Y \setminus S$ the covering space obtained by throwing away the branch points, consider the three groups:

$$\begin{aligned}
\mathrm{Aut}(X^\circ/Y^\circ) &= \{\text{continuous } \varphi\colon X^\circ \to X^\circ\colon p \circ \varphi = \varphi\};\\
\mathrm{Aut}(X/Y) &= \{\text{analytic } h\colon X \to X\colon f \circ h = h\};\\
\mathrm{Aut}(M(X)/\mathbb{C}(z)) &= \{\text{field homomorphisms } \vartheta\colon M(X) \to M(X)\colon \vartheta \text{ is}\\
&\qquad \text{the identity on } \mathbb{C}(z)\}.
\end{aligned}$$

The first is topological, the second analytic, the third algebraic. We claim that they are the same, after reversing the order of multiplication in the third group:

$$\mathrm{Aut}(X^\circ/Y^\circ) \cong \mathrm{Aut}(X/Y) \cong \mathrm{Aut}(M(X)/\mathbb{C}(z))^{\mathrm{opp}}.$$

Exercise 20.10. Prove this by showing that every deck transformation φ extends uniquely to an analytic isomorphism h, and showing that every automorphism of $M(X)$ that is the identity on $\mathbb{C}(z)$ has the form $f \mapsto f \circ h$ for a unique h. If you know some Galois theory, show that $M(X)$ is a Galois extension of $\mathbb{C}(z)$ if and only if the covering X° of Y° is a regular covering. Show that the Riemann surfaces of two algebraic curves are isomorphic if and only if their fields of meromorphic functions are isomorphic $\mathbb{C}$-algebras. Show that for any finite group G there is a Galois extension L of $\mathbb{C}(z)$ whose Galois group is G.

In fact, for a given compact Riemann surface Y, to give a compact Riemann surface X with a nonconstant analytic map from X to Y is equivalent to giving a finite-sheeted topological covering $X^\circ \to Y^\circ$ of the complement of a finite set, or to specify a finite extension $M(X)$ of the field $M(Y)$ of meromorphic functions. In this setting, the similarity seen in Proposition 13.23 between coverings and field extensions is more than just an analogy.

In most expositions the Riemann surface of a polynomial is constructed, following Weierstrass, by starting with a germ of an analytic function $w(z)$ satisfying the equation $F(z, w(z)) \equiv 0$, and analytically

continuing it around the plane. This approach, however, does not take advantage of the fact that the algebraic curve C is already, except for the modification and addition of a finite number of points, the desired Riemann surface.

Problem 20.11. Carry out this construction, using the ideas of §16b, and show that it gives the same Riemann surface.

20c. Holomorphic and Meromorphic 1-Forms

For any differentiable surface, as in Chapter 9, we can consider not just real differentiable 1-forms but complex ones as well, where a complex 1-form is given by $\omega_1 + i\omega_2$, with ω_1 and ω_2 real 1-forms. Again we can consider closed and exact forms, with $\omega_1 + i\omega_2$ being closed (resp. exact) when each of ω_1 and ω_2 is closed (resp. exact). The corresponding group of closed complex 1-forms modulo exact complex 1-forms is denoted $H^1(X;\mathbb{C})$. It is a complex vector space, which can be identified with $H^1(X)\oplus iH^1(X)$. If X is compact of genus g_X, $H^1(X;\mathbb{C})$ is a complex vector space of dimension $2g_X$.

When X is a compact Riemann surface, and not just a differentiable surface, there are some special closed complex 1-forms, called *holomorphic* 1-forms. They are the 1-forms that in local coordinates $\varphi_\alpha\colon U_\alpha \to X$, with $U_\alpha \subset \mathbb{C}$, have the form

$$f_\alpha\, dz = (u_\alpha + iv_\alpha)(dx + i\,dy) = (u_\alpha\, dx - v_\alpha\, dy) + i(v_\alpha\, dx + u_\alpha\, dy),$$

where $f_\alpha(z) = f_\alpha(x+iy) = u_\alpha(x,y) + iv_\alpha(x,y)$ is an analytic function. To define a global 1-form, using the notation of §19a, we must have

$$f_\alpha(z) \;=\; f_\beta(\varphi_{\beta\alpha}(z))\cdot\varphi_{\beta\alpha}'(z) \quad \text{on } U_{\beta\alpha}, \qquad \text{where} \quad \varphi_{\beta\alpha}' \;=\; \frac{d\varphi_{\beta\alpha}}{dz}$$

is the complex derivative. As before, the Cauchy–Riemann equations $\partial u_\alpha/\partial x = \partial v_\alpha/\partial y$ and $\partial u_\alpha/\partial y = -\partial v_\alpha/\partial x$ say that such a 1-form is closed.

The holomorphic 1-forms form a complex vector space, sometimes denoted $\Omega^{1,0}(X)$, or just $\Omega^{1,0}$. A holomorphic 1-form ω is exact precisely when there is an analytic function g on X with $dg = \omega$. In particular, if X is compact, every analytic function is constant, and so if ω is exact, then $\omega = 0$. This means that the natural map from $\Omega^{1,0}$ to $H^1(X;\mathbb{C})$ is injective. We regard $\Omega^{1,0}$ as a complex subspace of $H^1(X;\mathbb{C})$.

There is also a notion of an *antiholomorphic* 1-form. This is a 1-form that locally has the form

$$\overline{f_\alpha}\,\overline{dz} \;=\; (u_\alpha - iv_\alpha)(dx - i\,dy) = (u_\alpha\,dx - v_\alpha\,dy) + i(-v_\alpha\,dx - u_\alpha\,dy)$$

with $f_\alpha = u_\alpha + iv_\alpha$ an analytic function as above. Again these form a complex vector space, denoted $\Omega^{0,1}(X)$ or $\Omega^{0,1}$, and again these are closed forms, and the only antiholomorphic forms which are exact are differentials of complex conjugates $\overline{g}$ of complex analytic functions. Again, for X compact, we regard $\Omega^{0,1}$ as a subspace of $H^1(X;\mathbb{C})$.

It follows readily from the definitions that no nonzero 1-form can be both holomorphic and antiholomorphic: $\Omega^{1,0}\cap\Omega^{0,1}=0$. For any surface X the space $H^1(X;\mathbb{C})$ has a *complex conjugation* operator, that takes $\omega = \omega_1 + i\omega_2$ to $\overline{\omega} = \omega_1 - i\omega_2$. This is not a complex linear map, but is *conjugate linear*: it is linear as a map of real spaces, and for a complex number c, $\overline{c\omega} = \overline{c}\overline{\omega}$. Complex conjugation takes $\Omega^{1,0}$ to $\Omega^{0,1}$ and $\Omega^{0,1}$ to $\Omega^{1,0}$. The following is one of the major "existence theorems" about compact Riemann surfaces:

Theorem. *For a compact Riemann surface X, $H^1(X;\mathbb{C}) = \Omega^{1,0}\oplus\Omega^{0,1}$. Equivalently, $dim(\Omega^{1,0}) = dim(\Omega^{0,1}) = g_X$.*

In light of the preceding paragraph, the theorem is equivalent to showing that there are g_X linearly independent holomorphic 1-forms on X. We will say nothing about the proof of this theorem for a general compact Riemann surface, except to say that producing such 1-forms is closely related to producing meromorphic functions, which we have already discussed. For example, if ω_1 and ω_2 are two independent holomorphic 1-forms, then $\omega_2 = f\cdot\omega_1$, where f is a nonconstant meromorphic function. When the Riemann surface comes from an algebraic curve, however, we will prove this theorem in the next chapter. If the curve has a sufficiently nice form, however, it can be proved directly, as in the following problems:

Problem 20.12. Suppose $F(Z, W) = \Sigma_{i=0}^{n} a_i(Z)W^{n-i}$ is the polynomial as in §20a, and assume: (i) the curve C is nonsingular, i.e., there are no points (z_0, w_0) at which F, F_w, and F_z all vanish; and (ii) the degree of $a_i(Z)$ as a polynomial in Z is at most i, and if λ_i is the coefficient of Z^i in a_i, the equation $\Sigma_{i=0}^{n}\lambda_i t^{n-i} = 0$ has n distinct roots. (a) Show that X has n points over $\infty\in S^2$. (b) Show that $h = F_w(z, w)$ has a pole of order $n-1$ at each of these n points. (c) Show that the genus of

X is $(n-1)(n-2)/2$. (d) Show that

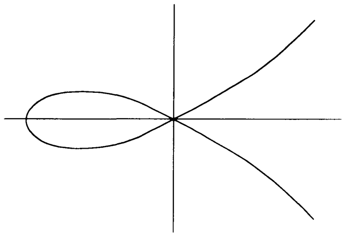
$$\Omega^{1,0}(X) = \left\{ \frac{g(z,w)}{F_w(z,w)} dz\colon g \text{ is a polynomial of degree at most } n-3 \text{ in } z \text{ and } w \right\},$$

which is a complex vector space of dimension g_X.

The following problem generalizes this to allow the curve C to have the simplest possible singularities: simple *nodes*. A node is a point where the two partial derivatives vanish, but the Hessian

$$\left(\frac{\partial^2 F}{\partial z\, \partial w}\right)^2 - \frac{\partial^2 F}{\partial z^2} \cdot \frac{\partial^2 F}{\partial w^2}$$

is not zero. Geometrically, this means the curve C has two nonsingular branches that cross transversely; in particular, there are two points of the Riemann surface over a node of C. The corresponding picture of a real curve ($W^2 = Z^2 + Z^3$) is:



Problem 20.13. Generalize the preceding problem to allow C to have a certain number δ of nodes, but no other singularities, but continue to assume (ii). Show that the genus of X is $(n-1)(n-2)/2 - \delta$, and that the holomorphic 1-forms on X are exactly those that have the form $(g(z,w)/F_w(z,w))\,dz$ where g is a polynomial of degree at most $n-3$ in z and w that vanishes at the nodes of C.

It is an important fact from the theory of algebraic curves that, after suitable algebraic transformations, every algebraic curve can be put in the form considered in the preceding problem. Note that it is necessary to allow singularities, since not every genus has the form $(n-1)(n-2)/2$. For a nice discussion of this, see Griffiths (1989).

One can also consider *meromorphic* 1-forms on X, which can be defined as for holomorphic 1-forms, locally given by $f_\alpha dz$, but with f_α only required to be meromorphic. If ω is a meromorphic 1-form, the order of ω at a point P, denoted $\mathrm{ord}_P(\omega)$, is defined to be the order of vanishing of f_α at the corresponding point in a coordinate disk. For example, when $X = S^2$, and $\omega = dz = -(z')^{-2}dz'$, with $z' = 1/z$, then ω has a pole of order 2 at the point at infinity. In general, we have

Proposition 20.14. *For any nonzero meromorphic 1-form ω on a compact Riemann surface X,*

$$\sum_{P\in X} \mathrm{ord}_P(\omega) \;=\; 2g_X - 2.$$

Proof. It is enough to prove the formula for one such ω, since any other has the form $h \cdot \omega$ for some meromorphic function h, and

$$\sum_{P\in X} \mathrm{ord}_P(h\cdot\omega) \;=\; \sum_{P\in X} \mathrm{ord}_P(h) + \mathrm{ord}_P(\omega) \;=\; 0 + \sum_{P\in X} \mathrm{ord}_P(\omega)$$

by Corollary 19.5. Assume that X comes equipped with a nonconstant meromorphic function $f\colon X \to S^2$. We take $\omega = df = f^*(dz)$. Near a point P of X, where the mapping in local coordinates is $t \mapsto t^e$, a meromorphic form on S^2 with local expression $g(z)\,dz$ pulls back to one on X with local expression $g(t^e)et^{e-1}\,dt$. Therefore $\mathrm{ord}_P(\omega) = (e_f(P) - 1) + e_f(P)\,\mathrm{ord}_{f(P)}(dz)$, and since the sum of $e_f(P)$ for P mapping to the point at infinity is n, and $\mathrm{ord}_\infty(dz) = -2$,

$$\sum_{P\in X} \mathrm{ord}_P(\omega) \;=\; \sum_{P\in X} (e_f(P) - 1) + n(-2);$$

this is $2g_X - 2$ by the Riemann–Hurwitz formula. □

Another proof can be given by appealing to the result of Chapter 8. Given a meromorphic 1-form ω with local expression $f_\alpha dz$, define a vector field V whose expression in the same local coordinates is given by $V_\alpha = 1/f_\alpha$, i.e., if $f_\alpha = u_\alpha + iv_\alpha$, then

$$V_\alpha \;=\; \left(\frac{u_\alpha}{{u_\alpha}^2 + {v_\alpha}^2}, \frac{-v_\alpha}{{u_\alpha}^2 + {v_\alpha}^2}\right).$$

Exercise 20.15. Verify that these V_α define a vector field V on X, and that $\mathrm{Index}_P V = -\mathrm{ord}_P(\omega)$.

Proposition 20.14 therefore also follows from Theorem 8.3.

Exercise 20.16. Reverse the argument in the above proof to give another proof of the Riemann–Hurwitz formula for $f: X \to Y$, under the assumption that Y has a nonzero meromorphic 1-form.

Define the *residue* of a meromorphic 1-form ω at a point P on a Riemann surface X, denoted $\mathrm{Res}_P\, \omega$, to be $(1/2\pi i)\int_\gamma \omega$, where γ is a small counterclockwise circle around P not surrounding any point except P where ω is not holomorphic.

Exercise 20.17. (a) Show that this is well defined. (b) If z is a local coordinate at P, and $\omega = f(z)\,dz$, with $f(z) = \sum_{n=-m}^{\infty} a_n z^n$, show that $\mathrm{Res}_P(\omega) = a_{-1}$.

Proposition 20.18 (Residue Formula). *If ω is a meromorphic 1-form on a compact Riemann surface X, then*

$$\sum_{P \in X} \mathrm{Res}_P(\omega) = 0.$$

Proof. We know that we can realize X as a polygon Π with sides identified. From the construction we see that the map from Π to X can be taken to be differentiable, and, moving the sides slightly, we may assume the image of each side is disjoint from the set of poles of ω. The 1-form ω then determines a closed 1-form ω' on Π, and we must show that the sum of the integrals of ω' around small circles around the poles of ω' is zero. Corollary 9.12 implies that this sum is the same as the integral of ω' around the boundary of Π, and this integral vanishes since the integrals over the sides that get identified in X cancel in pairs. □

The following problem gives another proof for Riemann surfaces coming from algebraic curves:

Problem 20.19. (a) Prove the Residue Formula directly when $X = S^2$. (b) With $z: X \to S^2$ as in §20b, and $\omega = f\,dz$, with $f \in M(X)$, show that, for $Q \in S^2$,

$$\sum_{P \in z^{-1}(Q)} \mathrm{Res}_P(f\,dz) = \mathrm{Res}_Q(g\,dz),$$

where g in $\mathbb{C}(z)$ is the trace of the $\mathbb{C}(z)$-linear endomorphism of $M(X)$ that is left multiplication by f. (c) Deduce the Residue Formula for $\omega = f\,dz$ from (a) and (b).

20d. Riemann's Bilinear Relations and the Jacobian

We have seen that the space $\Omega^{1,0}(X)$ of holomorphic 1-forms on X is a subspace of the De Rham group $H^1(X;\mathbb{C})$. The way $\Omega^{1,0}(X)$ sits in $H^1(X;\mathbb{C})$ is important both in studying functions on X and in studying the moduli of Riemann surfaces of given genus. Here we indicate how some of this is related to the facts about homology that we proved in Chapter 18. For this we use the pairing $(\omega, \nu) = \iint_X \omega \wedge \nu$ defined on closed 1-forms and their cohomology classes, but extended linearly to those with complex coefficients as usual. When applied to holomorphic 1-forms, there are two simple consequences of the definition:

(i) $(\omega, \nu) = 0$ if ω and ν are holomorphic; and
(ii) $i \cdot (\omega, \overline{\omega}) > 0$ if ω is nonzero and holomorphic.

The first follows from the fact that $dz \wedge dz = 0$. For the second, if, in local coordinates, $\omega = f(z)\, dz = f(z)(dx + i\,dy)$, then

$$i \cdot \omega \wedge \overline{\omega} \;=\; i \cdot |f(z)|^2 (dx + i\,dy) \wedge (dx - i\,dy) \;=\; 2|f(z)|^2\, dx \wedge dy,$$

which is strictly positive wherever f is not zero, so its integral is positive.

Taking a basis $a_1, \ldots, a_g, b_1, \ldots, b_g$ for homology as in Chapter 18, and applying Exercise 18.8, we deduce, for ω and ν holomorphic as above:

Proposition 20.20 (Riemann's Bilinear Relations).

(1) $$\sum_{j=1}^{g} \int_{a_j} \omega \int_{b_j} \nu \;=\; \sum_{j=1}^{g} \int_{a_j} \nu \int_{b_j} \omega;$$

(2) $$i \cdot \sum_{j=1}^{g} \left(\int_{a_j} \omega \int_{b_j} \overline{\omega} - \int_{a_j} \overline{\omega} \int_{b_j} \omega \right) \;>\; 0.$$

Corollary 20.21. *There is a unique basis $\omega_1, \ldots, \omega_g$ for the space of holomorphic 1-forms so that*

$$\int_{a_j} \omega_k \;=\; \begin{cases} 1 & \textit{if } j = k, \\ 0 & \textit{otherwise}. \end{cases}$$

Proof. This follows immediately from the fact that $H^1(X;\mathbb{C})$ is the direct sum of $\Omega^{1,0}(X)$ and $\Omega^{0,1}(X)$, together with the fact that integrating over cycles gives an isomorphism $H^1(X;\mathbb{C}) \cong \mathrm{Hom}(H_1X, \mathbb{C})$. □

The integrals of this basis over the other basis elements $b_1, \ldots, b_g$ carry important information. Let

$$\tau_{j,k} = \int_{b_k} \omega_j, \qquad 1 \leq j, k \leq g.$$

This gives a $(g \times g)$-matrix of complex numbers $Z = (\tau_{j,k})$, called the (normalized) *period matrix* corresponding to the homology basis. This matrix Z is far from arbitrary. It follows from what we have just seen that Z is nonsingular. Moreover,

Corollary 20.22. (*i*) *The matrix Z is symmetric, i.e.,* $\tau_{k,j} = \tau_{j,k}$. (*ii*) *The matrix Im(Z) whose entries are the imaginary parts of the entries of Z is a positive-definite symmetric matrix.*

Proof. (i) is an immediate consequence of the first of Riemann's bilinear relations, applied to the forms ω_j and ω_k.

(ii) follows from the second, applied to a form $\omega = \Sigma_{j=1}^{g} t_j \omega_j$, with t_j arbitrary real numbers, not all zero:

$$0 < i \cdot (\omega, \overline{\omega}) = i \cdot \sum_{j,k} t_j t_k (\overline{\tau}_{k,j} - \tau_{j,k}) = 2 \sum_{j,k} t_j t_k (\mathrm{Im}(\tau_{j,k}). \qquad \square$$

Of course, the period matrix is not unique, depending as it does on the choice of a homology basis, but if another basis is chosen, with the same intersection numbers, they will differ by a nonsingular matrix with integral entries that preserves this intersection pairing.

As in the plane, integrals between two points on X are determined up to these periods. Periods also play a role in a fundamental theorem of Abel, which concerns the question of when one can find a meromorphic function on a given Riemann surface with zeros and poles of given orders at given points.

Exercise 20.23. Show that for $X = S^2$, for any finite set of points P_i and any given integers m_i, provided $\Sigma m_i = 0$, there is a meromorphic function f on X with $\mathrm{ord}_{P_i}(f) = m_i$ for all i.

For $g > 0$, taking any basis $\omega_1, \ldots, \omega_g$ of the holomorphic 1-forms, we have a mapping from the homology group H_1X to $\mathbb{C}^g$ given by

$$\gamma \mapsto \left(\int_\gamma \omega_1, \ldots, \int_\gamma \omega_g\right).$$

It follows from what we have seen above that this map embeds H_1X in $\mathbb{C}^g$ as a *lattice*, i.e., the image of the basis elements $a_1, \ldots, b_g$

form a basis for $\mathbb{C}^g$ as a real vector space. If this lattice is denoted by Λ, then the quotient space (and group) $\mathbb{C}^g/\Lambda$ is called the *Jacobian* of X, and denoted $J(X)$; it is homeomorphic to $\mathbb{R}^{2g}/\mathbb{Z}^{2g}$, so to the Cartesian product of $2g$ circles.

If P and Q are any two points in X, they determine a point denoted $[Q-P]$ in $J(X)$, by the formula

$$[Q-P] = \left(\int_P^Q \omega_1, \ldots, \int_P^Q \omega_g\right) \in \mathbb{C}^g/\Lambda = J(X),$$

where the notation means to integrate along any path from P to Q; the resulting vector is defined up to an element in Λ. Similarly, given any 0-cycle $D = \Sigma m_i P_i$ of degree zero on X, one can define a point $[D]$ in the Jacobian by writing $D = \Sigma(Q_j - P_j)$ and setting $[D]$ equal to $\Sigma[Q_j - P_j]$. Equivalently, fix a point P_0 in X, and define the Abel–Jacobi mapping $A\colon X \to J(X)$ by the formula $A(P) = [P - P_0]$. Then $[\Sigma m_i P_i] = \Sigma m_i A(P_i)$.

Exercise 20.24. Show that this gives a well-defined homomorphism from the group $\tilde{Z}_0 X$ of 0-cycles of degree zero on X to $J(X)$.

The map from H_1X to $\mathbb{C}^g$ can be defined intrinsically, without choosing a basis of holomorphic 1-forms, as the map

$$H_1X \to \Omega^{1,0}(X)^*, \qquad \gamma \mapsto \left[\omega \mapsto \int_\gamma \omega\right],$$

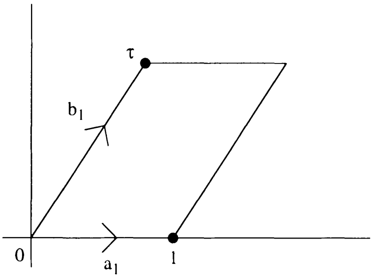
where $\Omega^{1,0}(X)^*$ is the dual space of complex-valued functions on $\Omega^{1,0}(X)$. This gives $J(X) = \Omega^{1,0}(X)^*/H_1(X)$, without choices.

The *divisor* $\mathrm{Div}(f)$ of a meromorphic function f is the 0-cycle $\Sigma\, \mathrm{ord}_P(f)P$. In the next chapter we will prove *Abel's theorem* that a zero cycle $D = \Sigma m_i P_i$ is the divisor of a meromorphic function if and only if its degree Σm_i is zero and $[D]$ is zero in $J(X)$. Equivalently, D is the boundary of a 1-chain γ such that $\int_\gamma \omega = 0$ for all holomorphic 1-forms ω. We will also prove the *Jacobi inversion theorem* that the Abel–Jacobi map from $\tilde{Z}_0 X$ to $J(X)$ is surjective.

20e. Elliptic and Hyperelliptic Curves

Before turning to the general situation it may be helpful to work out the simplest nontrivial case in detail. Consider a Riemann surface of the form $\mathbb{C}/\Lambda$, where Λ is a lattice. By Exercise 19.17 we may choose

such a lattice to be generated by 1 and τ, where τ is in the upper half plane. Realizing X by identifying the sides of the parallelogram



the images of the indicated sides can be taken as the standard basis a_1 and b_1 for H_1X. Now X has a natural holomorphic 1-form ω, whose pull-back to $\mathbb{C}$ is just the 1-form dz, with z the standard coordinate on $\mathbb{C}$. The integral of ω along a_1 is clearly 1, and the integral of ω along b_1 is τ. (Note that the corresponding period matrix is $Z = (\tau)$, in agreement with Corollary 20.22.) The mapping from H_1X to $\mathbb{C} = \mathbb{C}^g$, $\gamma \mapsto \int_\gamma \omega$, has image Λ generated by 1 and τ. Fix the point P_0 in X to be the image of the origin in $\mathbb{C}$. The Abel–Jacobi mapping from X to $\mathbb{C}/\Lambda$ then takes a point P to $[P - P_0] = \int_{P_0}^P \omega$. By looking at this mapping on the parallelogram spanned by 1 and τ, we see that this mapping is an isomorphism, i.e., $A: X \xrightarrow{\cong} \mathbb{C}/\Lambda$.

Now suppose X is any compact Riemann surface of genus 1, which we assume to have a nonzero holomorphic 1-form ω. We have the mapping $H_1X \to \mathbb{C}$ taking a homology class γ to $\int_\gamma \omega$. The image is a lattice Λ in $\mathbb{C}$. If we take a basis a_1, b_1 for H_1X as usual, we can take ω so that $\int_{a_1} \omega = 1$, in which case Λ is generated by 1 and $\tau = \int_{b_1} \omega$. As a very special case of Corollary 20.22 it follows that τ is in the upper half-plane. Fix a point P_0 in X.

Proposition 20.25. *The Abel–Jacobi mapping*

$$X \to \mathbb{C}/\Lambda, \qquad P \mapsto [P - P_0] = \int_{P_0}^{P} \omega,$$

is an isomorphism of X with $\mathbb{C}/\Lambda$.

Proof. Since the Abel–Jacobi mapping is analytic and nonconstant, it is unramified by the Riemann–Hurwitz formula, and therefore a covering map. We know from Chapter 13 that all finite coverings of $\mathbb{C}/\Lambda$ are given by subgroups of the fundamental groups, which means that X has the form $\mathbb{C}/\Lambda'$, where $\Lambda' \subset \Lambda$ is a subgroup of finite index.

But we proved directly that for any X of this form, the Abel–Jacobi map is an isomorphism. □

If a Riemann surface X is a two-sheeted covering of S^2 with four branch points, then by the Riemann–Hurwitz formula it has genus 1. By changing the map by an automorphism of S^2 one can take these four branch points to be 0, 1, ∞, and another $\lambda \in \mathbb{C}$. It follows easily (see below for a more general situation) that X is the Riemann surface of the curve $W^2 = Z(Z-1)(Z-\lambda)$. Since dz/w is a holomorphic 1-form on this Riemann surface, we know from what we have just seen that if we fix P_0 on X, the map

$$P \mapsto \int_{P_0}^{P} \frac{dz}{w} = \int_{P_0}^{P} \frac{dz}{\sqrt{z(z-1)(z-\lambda)}}$$

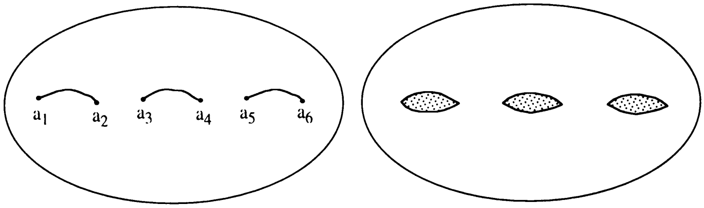
gives an isomorphism from X to $\mathbb{C}/\Lambda$. We will see that every Riemann surface of genus 1 is isomorphic to those arising this way, at least under the assumption that X comes from an algebraic curve. Most complex analysis texts have a chapter on such "elliptic integrals." (The Weierstrass $\wp$-function is used to find a two-sheeted branched covering from $\mathbb{C}/\Lambda$ to S^2.)

We end this brief excursion by looking at the special case of Riemann surfaces that can be realized as two-sheeted branched coverings of the sphere. By the Riemann–Hurwitz formula, such a covering must have an even number of branch points, namely, $2g_X + 2$. By Exercise 20.7, there is only one two-sheeted covering of S^2 with a given set of an even number of branch points. If these are the points $a_1, \ldots, a_m$ in $\mathbb{C}$, possibly together with the point ∞ (if m is odd), this can be realized as the Riemann surface of the algebraic curve

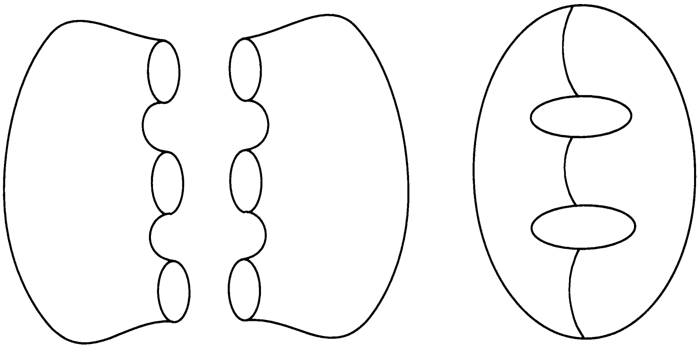
$$W^2 - \prod_{i=1}^{m} (Z - a_i) = 0.$$

These curves, and the corresponding Riemann surfaces, are called *hyperelliptic*. All Riemann surfaces of genus 2 are in fact hyperelliptic, but for genus greater than 2, not all Riemann surfaces arise this way.

The topology of a hyperelliptic surface can be seen directly, by cutting slits in the sphere along arcs from a_1 to a_2, a_3 to $a_4, \ldots, a_{2g+1}$ to a_{2g+2}.



The two-sheeted covering over the complement of these slits is trivial, so the Riemann surface can be constructed as in the picture:



Exercise 20.26. If $m = 2g_X + 2$, verify that the 1-forms $z^i\,dz/w$, for $0 \leq i \leq g - 1$, are holomorphic, and therefore give a basis for the holomorphic 1-forms.

CHAPTER 21

The Riemann–Roch Theorem

21a. Spaces of Functions and 1-Forms

Fix a compact Riemann surface X, and let $g = g_X$ be its genus, M its field of meromorphic functions, and Ω the space of meromorphic 1-forms on X. A *divisor* $D = \Sigma m_P P$ on X is just another word for a 0-chain. That is, it assigns an integer m_P to each point P in X, with only finitely many being nonzero. We say that the *order* of D at P is m_P, and write $\mathrm{ord}_P(D) = m_P$. The divisors on X form an abelian group. As for 0-chains, the *degree* of a divisor is the sum of the coefficients: $\deg(D) = \Sigma m_P$. If $E = \Sigma n_P P$ is another divisor, we write $E \geq D$ to mean that $n_P \geq m_P$ for all P in X. A divisor D is called *effective* if each coefficient m_P is nonnegative, i.e., $D \geq 0$.

Any nonzero meromorphic function f on X determines a divisor

$$\mathrm{Div}(f) = \Sigma\, \mathrm{ord}_P(f) P .$$

Similarly, any nonzero meromorphic 1-form ω on X determines a divisor

$$\mathrm{Div}(\omega) = \Sigma\, \mathrm{ord}_P(\omega) P .$$

Corollary 19.5 and Proposition 20.14 say that

$$\deg(\mathrm{Div}(f)) = 0 \quad \text{and} \quad \deg(\mathrm{Div}(\omega)) = 2g-2 .$$

Our goal in this chapter is to find meromorphic functions and 1-forms with prescribed, or at least controlled behavior. For example, we want to find functions with poles only at certain points, and with the orders

of poles at these points not exceeding some bounds. For a divisor $D = \Sigma m_P P$ on X let

$$\begin{aligned} L(D) &= \{f \in M\colon \operatorname{ord}_P(f) \geq -m_P \text{ for all } P \in X\} \\ &= \{f \in M\colon \operatorname{Div}(f) + D \geq 0\}. \end{aligned}$$

This set of functions $L(D)$ is a complex subspace of M. Similarly, let

$$\begin{aligned} \Omega(D) &= \{\omega \in \Omega\colon \operatorname{ord}_P(\omega) \geq m_P \text{ for all } P \in X\} \\ &= \{\omega \in \Omega\colon \operatorname{Div}(\omega) \geq D\}, \end{aligned}$$

a complex subspace of Ω. For example, $\Omega(0)$ is the space $\Omega^{1,0}$ of holomorphic 1-forms on X. Note that $L(D)$ allows poles at the points P where $m_P > 0$, while $\Omega(D)$ requires zeros at the same points.

Lemma 21.1. *(a) $L(D) = 0$ if $deg(D) < 0$, and $\Omega(D) = 0$ if $deg(D) > 2g - 2$.*

(b) For any D and any point Q in X, $L(D) \subset L(D+Q)$, and $\Omega(D) \supset \Omega(D+Q)$. In addition,

$$\dim(L(D+Q)/L(D)) \leq 1 \quad \textit{and} \quad \dim(\Omega(D)/\Omega(D+Q)) \leq 1\,.$$

(c) $L(D)$ and $\Omega(D)$ are finite-dimensional vector spaces.

Proof. (a) follows from the fact that $\deg(\operatorname{Div}(f)) = 0$ and $\deg(\operatorname{Div}(\omega)) = 2g - 2$. To prove (b), fix a local coordinate function z at Q, and let $m = \operatorname{ord}_Q(D)$. Any f in $L(D+Q)$ has a local expression $h(z)/z^{m+1}$, with h holomorphic at 0. The map which assigns $h(0)$ to f determines a homomorphism of complex vector spaces from $L(D+Q)$ to $\mathbb{C}$, whose kernel is exactly $L(D)$. This shows that either $L(D) = L(D+Q)$ or the quotient $L(D+Q)/L(D)$ is one dimensional. Similarly, any ω in $\Omega(D)$ has a local expression $h(z)z^m\,dz$ for h holomorphic, and assigning $h(0)$ to ω determines a map from $\Omega(D)$ to $\mathbb{C}$ whose kernel is $\Omega(D+Q)$.

It follows from (b) that $L(D)$ is finite dimensional if and only if $L(D+Q)$ is finite dimensional. Since one can get from any D to a divisor of negative degree by a finite number of subtractions of a point, the fact that $L(D)$ is finite dimensional follows from (a). Similarly for $\Omega(D)$, one can add points until the degree gets larger than $2g - 2$. □

Exercise 21.2. If $D \leq E$, show that

$$\dim(L(E)) - \dim(L(D)) \;\leq\; \deg(E) - \deg(D)\,.$$

One sees from the preceding proof that $\dim(L(D)) \leq \deg(D) + 1$.

For example, $L(0) = \mathbb{C}$ has dimension 1. If $D = Q$ is a point, however, we see from Exercise 20.7 that $L(Q)$ is also $\mathbb{C}$ unless $X \cong S^2$.

Exercise 21.3. If $X = S^2$, show that $\dim(L(D)) = \deg(D) + 1$ whenever $\deg(D) \geq -1$.

Lemma 21.4. (*a*) *For any nonzero meromorphic function f and any divisor D,*

$$\dim(L(D)) = \dim(L(D + \mathrm{Div}(f))).$$

(*b*) *For any nonzero meromorphic* 1*-form* ω *and any divisor D,*

$$\dim(\Omega(D)) = \dim(L(\mathrm{Div}(\omega) - D)).$$

Proof. We have isomorphisms

$$L(D) \to L(D + \mathrm{Div}(f)), \qquad h \mapsto h \cdot f,$$

and

$$L(\mathrm{Div}(\omega) - D) \to \Omega(D), \qquad h \mapsto h \cdot \omega,$$

from which the lemma follows. □

Although it has been fairly easy to get an upper bound for the size of $L(D)$, it is not so easy to get lower bounds, i.e., to show that there must be functions with given poles. When X comes from an algebraic curve, however, Proposition 20.8 gives a first step, for at least one divisor. Take $z\colon X \to S^2$ as in that proposition, and let E be the divisor of *poles* of z, that is,

$$E = \sum_{z(P)=\infty} e_z(P)P.$$

This is a divisor of degree n on X, where n is the degree of the mapping z.

Lemma 21.5. *For this divisor E, there is a constant k such that for all integers m,*

$$\dim(L(mE)) \geq \deg(mE) + 1 - k = mn + 1 - k.$$

Proof. We need the following fact: for any meromorphic function h on X, there is a nonzero polynomial $p(z)$ in $\mathbb{C}[z]$ and an integer t so that $p(z) \cdot h$ is in $L(tE)$. To prove this, we need only find $p(z)$ so that $p(z) \cdot h$ has no poles outside E, and one sees that $\prod(z - z(P))^{-\mathrm{ord}_P(h)}$, the product over all P such that $z(P) \neq \infty$ and $\mathrm{ord}_P(h) < 0$, is such a polynomial.

We saw in Proposition 20.8 (see Lemma C.19) that M is a vector space over $\mathbb{C}(z)$ of dimension n. By the fact proved in the preceding paragraph, we can find a basis $h_1, \ldots, h_n$ for M over $\mathbb{C}(z)$ and an integer t so that each h_i is in $L(tE)$. Now for $m = t + s$, $s \geq 0$, the $(s+1) \cdot n$ functions $z^j \cdot h_i$, $0 \leq j \leq s$, $1 \leq i \leq n$, are all in $L(mE)$. This means that, for such m, the dimension of $L(mE)$ is at least $(m - t + 1) \cdot n = mn + 1 - k$ for some constant k. Increasing k if necessary, one may also achieve this inequality for the finite number of m with $0 \leq m < t$. The inequality is automatic for $m < 0$ and any $k \geq 0$, so the lemma is proved. □

Lemma 21.6. *There are integers k and N so that*

$$\dim(L(D)) \geq \deg(D) + 1 - k$$

for all divisors D on X, with equality

$$\dim(L(D)) = \deg(D) + 1 - k \qquad \text{if} \quad \deg(D) \geq N.$$

Proof. Choose E as above, and define k to be the smallest integer so that Lemma 21.5 holds for k. Suppose that D is a divisor on X such that $D \leq mE$ for some integer m. It follows from Exercise 21.2 that

$$\dim(L(D)) \geq \deg(D) + \dim(L(mE)) - \deg(mE) \geq \deg(D) + 1 - k,$$

which proves the required inequality for such a divisor D. Given any divisor D on X, there is a nonzero meromorphic function h such that $D - \mathrm{Div}(h) \leq mE$ for some integer m. Indeed, as in the preceding lemma, one can take h to be $\Pi(z - z(P))^{\mathrm{ord}_P(D)}$, the product over all P with $z(P) \neq \infty$. Then by Lemma 21.4(a) and the result just proved,

$$\begin{aligned} \dim(L(D)) &= \dim(L(D - \mathrm{Div}(h))) \\ &\geq \deg(D - \mathrm{Div}(h)) + 1 - k = \deg(D) + 1 - k. \end{aligned}$$

By the minimality of k, there is some divisor D_0 such that the dimension of $L(D_0)$ is $\deg(D_0) + 1 - k$. From Exercise 21.2 it follows that for any divisor D such that $D \geq D_0$,

$$\dim(L(D)) \leq \deg(D) + \dim(L(D_0)) - \deg(D_0) \leq \deg(D) + 1 - k,$$

so $\dim(L(D)) = \deg(D) + 1 - k$ for any such D.

Let $N = \deg(D_0) + k$. If the degree of D is at least N, then the degree of $D - D_0$ is at least k, so the dimension of $L(D - D_0)$ is at least $k + 1 - k = 1$. There is therefore a nonzero function f in $L(D - D_0)$, which means that $D + \mathrm{Div}(f) \geq D_0$, and so

$$\begin{aligned} \dim(L(D)) &= \dim(L(D + \mathrm{Div}(f))) \\ &= \deg(D + \mathrm{Div}(f)) + 1 - k = \deg(D) + 1 - k, \end{aligned}$$

which proves the lemma. □

There can be only one integer k with the second property in the lemma, so k depends only on the Riemann surface X. We will see in §21c that k is the genus of X, and that N can be taken to be $2g-2$.

21b. Adeles

Lemma 21.1(b) says that each of the two subspaces $L(D) \subset L(D+Q)$ and $\Omega(D+Q) \subset \Omega(D)$ are either equalities or subspaces of codimension one. These cannot both be subspaces of codimension one, for if ω is in $\Omega(D)$ and f is in $L(D+Q)$, then $f \cdot \omega$ is a meromorphic 1-form with at most one simple pole at Q; the Residue Formula then implies that $\text{Res}_Q(f \cdot \omega) = 0$, which means that $f \cdot \omega$ does not have a pole at Q, and hence either ω is in $\Omega(D+Q)$ or f is in $L(D)$. We will eventually see that one of these inclusions is an equality exactly when the other is not, and this is the core of the proof of the Riemann–Roch theorem. What we will do in this section is to prove a kind of local version of this assertion.

For a point P in X, let us denote by M_P the *germs* of meromorphic functions at P. These germs are defined as in §16b, by taking equivalence classes of meromorphic functions in neighborhoods of P, two being equivalent if they agree on some (punctured) neighborhood of P. If z is a local coordinate at P, any such germ has a unique power series expansion $\Sigma_{n=-m}^{\infty} a_n z^n$. If f is in M_P, and ω is a meromorphic 1-form on X, the *residue* $\text{Res}_P(f \cdot \omega)$ can be defined to be $1/2\pi i$ times the integral of $f \cdot \omega$ around a small counterclockwise circle around P. In local coordinates, $f \cdot \omega$ can be written $\Sigma b_n z^n\, dz$, and this residue is b_{-1}.

Define an *adele* on X to be the assignment of a germ f_P of a meromorphic function at P for every point P in X, with the property that $\text{ord}_P(f_P) \geq 0$ (i.e., f_P is holomorphic at P) for all but finitely many P. We write $\mathbf{f} = (f_P)$ for the adele defined by such a collection of functions f_P. These adeles form a complex vector space, which we denote by R. Any meromorphic function f on X determines an adele, by assigning the germ of f at P to each P, so the field M of meromorphic functions is a subspace of R. An adele can be thought of as a kind of "discontinuous function" on X. Since there is no relation between the "values" f_P at different points of X, it is remarkable that they can be a useful tool.

If $\mathbf{f} = (f_P)$ is an adele, then $\text{Res}_P(f_P \cdot \omega) = 0$ for all but finitely many P (those where f_P or ω has a pole). We can therefore add the residues

$\mathrm{Res}_P(f_P \cdot \omega)$ over all P in X, getting a complex number. In other words, ω defines a homomorphism

$$\varphi_\omega : R \to \mathbb{C}, \qquad \mathbf{f} = (f_P) \mapsto \sum_{P \in X} \mathrm{Res}_P(f_P \cdot \omega),$$

which is a linear map of complex vector spaces. If $D = \Sigma m_P P$ is a divisor such that ω is in $\Omega(D)$, and $\mathrm{ord}_P(f_P) \geq -m_P$ for all P, then $f_P \cdot \omega$ is holomorphic at P, so the residue is zero. Define $R(D) \subset R$ by the formula

$$R(D) = \{\mathbf{f} = (f_P) \in R \colon \mathrm{ord}_P(f_P) \geq -\mathrm{ord}_P(D) \text{ for all } P \in X\}.$$

This means that the homomorphism φ_ω vanishes on $R(D)$. In addition, the Residue Formula says that if these f_P all come from one meromorphic function f on X, then $\Sigma_{P \in X} \mathrm{Res}_P(f \cdot \omega) = 0$, so φ_ω also vanishes on the subspace M of R. It follows that φ_ω determines a homomorphism (still denoted φ_ω)

$$\varphi_\omega : R/(R(D) + M) \to \mathbb{C}.$$

Define $S(D)$ to be this complex vector space $R/(R(D) + M)$, and define $\Omega'(D)$ to be the dual space

$$\Omega'(D) = S(D)^* = \mathrm{Hom}_{\mathbb{C}}(R/(R(D) + M), \mathbb{C}).$$

Then φ_ω is an element of this space $\Omega'(D)$. What we have done is construct a natural homomorphism from $\Omega(D)$ to $\Omega'(D)$, taking ω to φ_ω. In the next section we will show that this homomorphism is an isomorphism.

Exercise 21.7. Show that the homomorphism from $\Omega(D)$ to $\Omega'(D)$ is injective, and that a meromorphic 1-form ω is in $\Omega(D)$ if and only if φ_ω vanishes on $R(D)$.

In this section we prove that $\Omega'(D)$ has some properties we would like $\Omega(D)$ to have. As in the preceding section, the main idea is to compare $\Omega'(D)$ and $\Omega'(D + Q)$ for Q a point in X.

Since $R(D)$ is contained in $R(D + Q)$, there is a canonical surjection from $R/(R(D) + M)$ onto $R/(R(D + Q) + M)$, i.e., from $S(D)$ onto $S(D + Q)$. The kernel is $(R(D + Q) + M)/(R(D) + M)$.

Lemma 21.8. *For any divisor D and point Q,*

$$\dim((R(D + Q) + M)/(R(D) + M)) \leq 1,$$

with equality if and only if $L(D) = L(D + Q)$.

Proof. Choose a germ g_Q at Q such that $\mathrm{ord}_Q(g_Q) = -\mathrm{ord}_Q(D) - 1$, and let $g_P = 0$ for $P \neq Q$. The adele $\mathbf{g} = (g_P)$ gives a generator of the quotient space in the lemma. Indeed if $f = (f_P)$ is any adele in $R(D+Q)$, then there is some scalar λ so that $f - \lambda g$ is in $R(D)$. This element will be nonzero exactly when there is no h in M with $g - h$ in $R(D)$, which says exactly that there is no h in $L(D+Q)$ that is not in $L(D)$. □

Let k be the integer from Lemma 21.6.

Lemma 21.9.

(*a*) *If D is a divisor such that $\dim(L(D)) = \deg(D) + 1 - k$, then $R(D) + M = R$, so $S(D) = 0$ and $\Omega'(D) = 0$.*
(*b*) *For any D the space $S(D)$ has finite dimension, so its dual space $\Omega'(D)$ has the same finite dimension.*
(*c*) *For any divisor D and point Q, $\Omega'(D+Q)$ is a subspace of $\Omega'(D)$, and*

$$\dim(\Omega'(D)/\Omega'(D+Q)) \leq 1,$$

with equality if and only if $L(D) = L(D+Q)$.
(*d*) *For any nonzero meromorphic function f on X,*

$$\dim(\Omega'(D + \mathrm{Div}(f))) = \dim(\Omega'(D)).$$

Proof. If $\dim(L(D)) = \deg(D) + 1 - k$, then for any point Q we know that $\dim(L(D+Q)) = \deg(D+Q) + 1 - k$ (see Exercise 21.2), i.e., $L(D+Q) \neq L(D)$. By the preceding lemma, this means that $R(D+Q) \subset R(D) + M$. Continuing to add points to $D+Q$, we see that $R(E) \subset R(D) + M$ for all divisors E such that $E \geq D$. But any element of R is in $R(E)$ for some such E, so $R = R(D) + M$, which proves (a).

For (b), take a sequence of surjections $S(D) \twoheadrightarrow S(D+Q_1) \twoheadrightarrow S(D+Q_1+Q_2) \twoheadrightarrow \ldots \twoheadrightarrow S(E)$, until E is large enough so (a) implies that $S(E)$ is zero. Lemma 21.8 implies that the kernel of each of these surjections is at most one dimensional, so by induction each $S(D)$ must be finite dimensional.

Dual to the exact sequence

$$0 \to (L(D+Q)+M)/(L(D)+M) \to S(D) \to S(D+Q) \to 0$$

is the exact sequence

$$0 \to \Omega'(D+Q) \to \Omega'(D) \to ((L(D+Q)+M)/(L(D)+M))^* \to 0.$$

This shows that the inclusion $\Omega'(D+Q) \hookrightarrow \Omega'(D)$ is either an iso-

morphism or its cokernel has dimension one. By the preceding lemma, we see that the latter occurs exactly when $L(D) = L(D+Q)$, which proves (c).

For (d), there is a natural isomorphism from $R(D + \mathrm{Div}(f))$ to $R(D)$ that takes $\mathbf{f}$ to $f \cdot \mathbf{f}$. This determines an isomorphism from $S(D + \mathrm{Div}(f))$ to $S(D)$, and, taking duals, from $\Omega'(D)$ to $\Omega'(D + \mathrm{Div}(f))$. □

Lemma 21.10. *For any divisor D on X,*

$$\dim(L(D)) = \deg(D) + 1 - k + \dim(\Omega'(D)).$$

Proof. This equation is certainly true if $\deg(D) \geq N$, with N as in Lemma 21.6, for then $\dim(L(D)) = \deg(D) + 1 - k$ and $\dim(\Omega'(D)) = 0$ by Lemma 21.6 and Lemma 21.9(a). Since we can get between any two divisors by successively adding and subtracting points, it suffices to show that the equation is true for a divisor D if and only if it is true for $D + Q$, where Q is any point. Comparing the two equations, what must be proved is that

$$\dim(L(D+Q)) - \dim(L(D)) + \dim(\Omega'(D)) - \dim(\Omega'(D+Q)) = 1,$$

and this is simply a translation of (c) in the preceding lemma. □

Let Ω' be the union of all $\Omega'(D)$, taken over all divisors D. An element of Ω' is a homomorphism from R to $\mathbb{C}$ which vanishes on M and vanishes on some (unspecified) $R(D)$. The space Ω of meromorphic differentials on X maps to Ω', and we want to see that this is an isomorphism. We have seen that if ω is any nonzero meromorphic differential, any other can be written in the form $f \cdot \omega$ for some meromorphic function f. This means that Ω is a one-dimensional vector space over the field M. The space Ω' is also a vector space over M, by the rule that if f is in M and $\varphi \colon R \to \mathbb{C}$, then $f \cdot \varphi$ is the homomorphism which takes $\mathbf{f}$ to $\varphi(f \cdot \mathbf{f})$.

Lemma 21.11. *The dimension of* Ω' *over M is* 1.

Proof. We know that Ω' is not zero, for example, by applying the preceding lemma for D of small degree to see that $\Omega'(D) \neq 0$. To complete the proof we must show that two elements φ and ψ in Ω' cannot be independent over M. If this were the case, then for any elements $h_1, \ldots, h_n$ of M which are independent over $\mathbb{C}$, it would follow that $h_1\varphi, \ldots, h_n\varphi, h_1\psi, \ldots, h_n\psi$ are elements of Ω' which are independent over $\mathbb{C}$. Take a divisor E which is large enough so that φ and ψ are in $\Omega'(E)$. Suppose the functions h_i are a basis for $L(D)$ for some D. Then the $2n$ products $h_i\varphi$ and $h_i\psi$ are in $\Omega'(E - D)$,

so

$$\dim(\Omega'(E-D)) \geq 2\dim(L(D)) \geq 2(\deg(D)+1-k).$$

By Lemma 21.10,

$$\dim(\Omega'(E-D)) = k-1-\deg(E-D)+\dim(L(E-D)).$$

Now $\deg(E-D) = \deg(E)-\deg(D)$, and $L(E-D)=0$ provided $\deg(E-D)<0$. So if we take any D with $\deg(D)>\deg(E)$, the displays lead to the inequality

$$2(\deg(D)+1-k) \leq \deg(D)-\deg(E)+k-1,$$

which says that $\deg(D) \leq 3k-3-\deg(E)$. But we may take D of arbitrarily large degree, which is a contradiction. □

The canonical homomorphism from the space Ω of meromorphic differentials to the space Ω' is a homomorphism of vector spaces over the field M. It is not identically zero by Exercise 21.7, since one can certainly find meromorphic differentials ω and adeles $\mathbf{f}$ such that there is exactly one point P at which $f_P \cdot \omega$ has a simple pole. But since both vector spaces have dimension one over M, it follows that the map $\Omega \rightarrow \Omega'$ is an isomorphism. We saw at the beginning that the subspace $\Omega(D)$ of Ω is mapped into $\Omega'(D)$ by this map, and it follows from Exercise 21.7 that $\Omega(D)$ maps isomorphically onto $\Omega'(D)$. So we have proved

Proposition 21.12. *The canonical map $\Omega(D) \rightarrow \Omega'(D)$ is an isomorphism.*

Exercise 21.13. Given germs $f_1, \ldots, f_n$ of meromorphic functions at distinct points $P_1, \ldots, P_n$ of X, and integers $m_1, \ldots, m_n$, show that there is a meromorphic function f so that $\mathrm{ord}_{P_i}(f-f_i) \geq m_i$ for $1 \leq i \leq n$.

21c. Riemann–Roch

From Lemma 21.10 and Proposition 21.12 we have the formula

$$(*) \qquad \dim(L(D)) = \deg(D)+1-k+\dim(\Omega(D))$$

for all divisors D on X, valid for some integer k which we do not yet know. We can specialize (*) to some cases where we know some-

thing. For example, if we take $D = 0$, $L(0) = \mathbb{C}$, and the formula says that $1 = 0 + 1 - k + \dim(\Omega(0))$, i.e., that the space of holomorphic 1-forms has dimension k. Fix any meromorphic divisor ω, and let K be the divisor of ω, which we know is a divisor of degree $2g-2$. Applying Lemma 21.4(b) with $D = K$, we get

$$\dim(\Omega(K)) = \dim(L(K-K)) = \dim(L(0)) = 1,$$

and applying the same lemma with $D = 0$, we get

$$\dim(L(K)) = \dim(\Omega(0)) = k.$$

Now apply (*) with $D = K$, yielding

$$k = (2g-2) + 1 - k + 1,$$

which means that k must be g. So we have proved:

Theorem 21.14 (Riemann–Roch Theorem). *If X is the Riemann surface of an algebraic curve, then for any divisor D on X,*

$$\begin{aligned}\dim(L(D)) &= \deg(D) + 1 - g + \dim(\Omega(D)) \\ &= \deg(D) + 1 - g + \dim(L(K-D)),\end{aligned}$$

where K is the divisor of any nonzero meromorphic 1-form on X.

Corollary 21.15. *The space of holomorphic differentials has dimension g.*

This proves that $H^1(X;\mathbb{C}) = \Omega^{1,0}(X) \oplus \Omega^{0,1}(X)$.

Corollary 21.16. *$\dim(L(D)) \geq \deg(D) + 1 - g$, with equality whenever $\deg(D) \geq 2g-1$.*

Corollary 21.17. *For any two points P and Q on X, there is a meromorphic 1-form φ with simple poles at P and Q and no other poles.*

Proof. Riemann–Roch for $D = -P - Q$ gives

$$0 = -2 + 1 - g + \dim(\Omega(-P-Q)),$$

or $\dim(\Omega(-P-Q)) = g + 1$. This means that there is a meromorphic 1-form φ that is in $\Omega(-P-Q)$ but not in $\Omega(0)$. Since the sum of the residues is zero, φ must have simple poles at both P and Q, with the residue at Q being minus the residue at P. □

Corollary 21.18. *If $g_X = 0$, then X is isomorphic to S^2.*

Proof. Take any point P. Since $\dim(L(P)) \geq \deg(P) + 1 - g = 2$, there

is a nonconstant meromorphic function f with at most one pole. This is a mapping from X to S^2 of degree 1, so is an isomorphism. □

Exercise 21.19. If $g_X = 1$, show that there is an analytic mapping $f: X \to S^2$ of degree 2. Deduce that X is the Riemann surface of a curve $W^2 = Z(Z-1)(Z-\lambda)$, $\lambda \neq 0, 1$.

Exercise 21.20. Assume that $g = g_X \geq 1$. (a) Show that there are g distinct points $P_1, \ldots, P_g$ on X so that $\Omega(P_1 + \ldots + P_g) = 0$. (b) Show that there are points $P_1, \ldots, P_g$ on X so that $\Omega(P_1 + \ldots + P_g) \neq 0$. (c) If $g \geq 2$, show that there is an analytic mapping $f: X \to S^2$ of degree at most g. In particular, if $g = 2$, X is hyperelliptic.

Exercise 21.21. Show that the 1-form φ of Corollary 21.17 is unique up to multiplying by a nonzero scalar and adding a holomorphic 1-form. Show that there is a unique such φ whose residue at P is 1, whose residue at Q is -1, and so that $\int_\gamma \varphi$ is purely imaginative for all closed paths γ in $X \setminus \{P, Q\}$.

Exercise 21.22. A real-valued function u on a Riemann surface is *harmonic* if it is locally the real part of an analytic function. A function which is harmonic in a punctured neighborhood of a point P is said to have a *logarithmic pole* at P if, with z a local coordinate at P, there is a nonzero real scalar a so that $u - a \cdot \log(|z|)$ extends to be harmonic in a neighborhood of P. Show that for any two points P and Q on X, there is a harmonic function u on $X \setminus \{P, Q\}$ that has logarithmic poles at P and Q.

Exercise 21.23. For any point P on X, show that there is a meromorphic differential φ on X with a double pole at P. Deduce that there is a harmonic function u on $X \setminus \{P\}$ so that, if $z = x + iy$ is a local coordinate at P, then $u - x/(x^2 + y^2)$ is harmonic near P.

Historically, the arguments went in the reverse order: harmonic functions were found with the properties in the preceding exercises, and these were used to find meromorphic 1-forms and to prove Riemann–Roch. By regarding harmonic functions as integrals of fluid flows or electric fields on X, one can give intuitive arguments for their existence, say by putting sources and sinks at the points P and Q. For a lively discussion along these lines, see Klein (1893).

Exercise 21.24. Given any sequence $P_1, P_2, \ldots, P_n, \ldots$ of points

in X, show that there are exactly g positive integers k, all in the interval $[1, 2g-1]$, such that

$$L(P_1 + \ldots + P_{k-1}) = L(P_1 + \ldots + P_k).$$

When all P_i are taken to be a fixed point P, these integers are called the *Weierstrass gaps at P*.

Exercise 21.25. Suppose D and E are divisors on X such that $D+E$ is the divisor of a meromorphic 1-form. Prove *Brill–Noether reciprocity:* $\dim(L(D)) - \dim(L(E)) = {}^1\!/\!_2(\deg(D) - \deg(E))$.

Exercise 21.26. If $z\colon X \to S^2$ is an analytic mapping of degree n, and Q in S^2 is a point such that $z^{-1}(Q) = \{P_1, \ldots, P_n\}$ has n distinct points, use Riemann–Roch to show that there is a meromorphic function w on X so that $w(P_1), \ldots, w(P_n)$ are distinct complex numbers. If $F(z, w) = 0$ is the irreducible equation for w over $\mathbb{C}(z)$ (with denominators cleared), show that X is isomorphic to the Riemann surface of F.

Once one knows Riemann–Roch for a general compact Riemann surface X, the preceding exercise shows that X comes from an algebraic curve.

Exercise 21.27. Show that the Riemann surface X of the polynomial $W^4 + Z^4 - 1 = 0$ has genus 3, but X is not hyperelliptic. Show in fact that $\dim\Omega(2P) = 1$ for all P in X.

21d. The Abel–Jacobi Theorem

In this section we prove the assertions made in §20d. The first is Abel's criterion for when a divisor $D = \Sigma m_i P_i$ is a divisor of a meromorphic function: it must have degree zero and be in the kernel of the Abel–Jacobi map. We use the notation of that section.

Theorem 21.28 (Abel's Theorem). *There is a meromorphic function f on X with* $\mathrm{Div}(f) = D$ *if and only if* $\deg(D) = 0$ *and* $[D] = 0$ *in* $J(X)$.

We first sketch the proof of the necessity of these conditions. Suppose f is a nonconstant meromorphic function on X, giving a mapping of degree n from X to S^2, with branch set S. Corollary 19.5 shows

that $\deg(\mathrm{Div}(f)) = 0$. Fix a point P_0 in X. Consider the mapping from $S^2 \setminus S$ to $J(X)$ that takes a point Q to the point $\Sigma_{i=1}^{n} [P_i - P_0]$, where $P_1 \ldots, P_n$ are the points of X in $f^{-1}(Q)$. It is not hard to see that this extends continuously to the branch points, giving a continuous mapping from S^2 to $J(X)$. Since S^2 is simply connected, by Proposition 13.5 this mapping factors: $S^2 \to \mathbb{C}^g \to \mathbb{C}^g/\Lambda$. By looking locally, one can verify that each of the g coordinate maps are analytic functions on S^2. But analytic functions on S^2 are constant, so the given map from S^2 to $J(X)$ must be constant. The fact that the value at 0 is equal to the value at ∞ is precisely the condition that $[\mathrm{Div}(f)] = [\Sigma \operatorname{ord}_P(f) \cdot P] = 0$ in $J(X)$.

We turn now to the converse. Let D be a divisor of degree zero in the kernel of the Abel–Jacobi map. We must show that D is the divisor of a meromorphic function. The motivation for the proof comes from the fact that if f is a meromorphic function on X, and we set $\varphi = df/f$, then φ is a meromorphic differential on X with at most simple poles and, in fact, for any P, $\mathrm{Res}_P(\varphi) = \operatorname{ord}_P(f)$. We will look for a meromorphic 1-form φ with at most simple poles among the points appearing in D, such that $\mathrm{Res}_P(\varphi) = \operatorname{ord}_P(D)$ for all P. Then we will define a function f on X by the formula $f(P) = \exp(\int_{P_0}^{P} \varphi)$, where P_0 is a fixed point. Provided this is well defined, it will satisfy the equation $\varphi = df/f$, and so we will have $\operatorname{ord}_P(f) = \operatorname{ord}_P(D)$ for all P, so $\mathrm{Div}(f) = D$.

Since the degree of D is zero, we may write $D = \Sigma_{i=1}^{r}(P_i - Q_i)$, for some points $P_1, \ldots, Q_r$ (not necessarily unique). Let $S = \{P_1, \ldots, P_r, Q_1, \ldots, Q_r\}$. We know by Corollary 21.17 that there is a meromorphic 1-form φ_i with simple poles at P_i and Q_i (only), and with residues 1 at P_i and -1 at Q_i. Let $\varphi = \Sigma_{i=1}^{r} \varphi_i$. We want to define $f(P) = \exp(\int_{P_0}^{P} \varphi)$, where the integral is along any path from P_0 to P in $X \setminus S$. This will be well defined provided the integral of φ along any closed path τ in $X \setminus S$ is in $2\pi i\mathbb{Z}$. The form φ is only defined up to the addition of a holomorphic 1-form, so the proof of the Abel–Jacobi theorem is reduced to the

Claim 21.29. *There is a holomorphic* 1*-form* ω *so that* $\int_\tau(\varphi - \omega)$ *is in* $2\pi i\mathbb{Z}$ *for all* 1*-cycles* τ *on* $X \setminus S$.

We need the following refinement of Exercise 18.8. We take $2g$ closed arcs a_j and b_j as in §17c. Cutting the surface open along these arcs, we realize it as the identification space of a plane polygon Π with sides identified. These choices can be made so that the map from Π to X is a diffeomorphism on the interior of Π, and has a $\mathscr{C}^\infty$ exten-

sion to a neighborhood of Π. By means of this map 1-forms on X correspond to 1-forms on Π. Let ω be a closed 1-form on X, and define a function h on the closure of Π by the formula $h(P) = \int_{P_0}^{P} \omega$, for some fixed point P_0 in Π. Let φ be a closed 1-form defined on a neighborhood of the union of these $2g$ arcs in X, so φ determines a 1-form on a neighborhood of the boundary $\partial\Pi$ of Π.

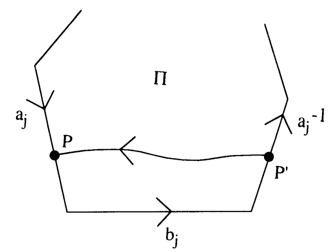
Lemma 21.30.

$$\int_{\partial\pi} h\varphi = \sum_{j=1}^{g}\left(\int_{a_j}\omega\int_{b_j}\varphi - \int_{a_j}\varphi\int_{b_j}\omega\right).$$

Proof. Note first that if P and P' are corresponding points of the boundary edges a_j and a_j^{-1} of Π, then

$$h(P) - h(P') = -\int_{b_j}\omega,$$

as one sees by integrating along a path from P' to P, noting that the integrals over corresponding parts of a_j and a_j^{-1} cancel.



Therefore

$$\int_{a_j} h\varphi + \int_{a_j^{-1}} h\varphi = \int_{a_j}(h(P) - h(P'))\,\varphi(P)$$
$$= \int_{a_j}\left(-\int_{b_j}\omega\right)\varphi = -\int_{b_j}\omega\cdot\int_{a_j}\varphi.$$

Similarly if Q and Q' are corresponding points of b_j and b_j^{-1}, then

$$h(Q) - h(Q') = \int_{a_j}\omega,$$

so

$$\int_{b_j} h\varphi + \int_{b_j^{-1}} h\varphi = \int_{b_j}(h(Q) - h(Q'))\,\varphi(Q) = \int_{a_j}\omega\cdot\int_{b_j}\varphi.$$

Adding over all the edges of the boundary of Π gives the identity of the lemma. □

To apply the lemma in our situation, the arcs a_j, b_j must be taken so that none of them goes through a point of S. Let σ_i be a path from Q_i to P_i which does not hit any of the arcs a_j or b_j.

Lemma 21.31. *For any holomorphic 1-form ω on X,*

$$(2\pi i)\sum_{i=1}^{r}\int_{\sigma_i}\omega = \sum_{j=1}^{g}\left(\int_{a_j}\omega\int_{b_j}\varphi - \int_{a_j}\varphi\int_{b_j}\omega\right).$$

Proof. We apply the preceding lemma. We must evaluate $\int_{\partial\Pi} h\varphi$, with $h(P) = \int_{P_0}^{P}\omega$. The Residue Formula in the polygon Π gives

$$\int_{\partial\Pi} h\varphi = (2\pi i)\sum_{p\in\Pi}\mathrm{Res}_P(h\varphi) = (2\pi i)\left(\sum_{i=1}^{r} h(P_i) - \sum_{i=1}^{r} h(Q_i)\right),$$

the last since φ_i has simple poles, with residues 1 and -1 at P_i and Q_i. And

$$\sum_{i=1}^{r} h(P_i) - \sum_{i=1}^{r} h(Q_i) = \sum_{i=1}^{r}\int_{\sigma_i}\omega$$

by definition. So Lemma 21.30 gives the required conclusion. □

The hypothesis that D is in the kernel of the Abel–Jacobi map means that there is a 1-cycle γ such that $\Sigma_{i=1}^{r}\int_{\sigma_i}\omega = \int_{\gamma}\omega$ for all holomorphic 1-forms ω. So we have, for all holomorphic ω,

$$(2\pi i)\int_{\gamma}\omega = \sum_{j=1}^{g}\left(\int_{a_j}\omega\int_{b_j}\varphi - \int_{a_j}\varphi\int_{b_j}\omega\right).$$

We can now prove Claim 21.29. Since a basis of $H_1(X \setminus S)$ is given by the cycles a_j, b_j, and small circles around the points in S (either by Mayer–Vietoris or Problem 17.12), it suffices to find a holomorphic 1-form ω such that the integral of $\varphi - \omega$ around all these cycles is in $2\pi i\mathbb{Z}$. Note that for any such φ, the integral around a small circle around a point in S is in $2\pi i\mathbb{Z}$, since all the residues of φ are integers. To start, let $\lambda_j = \int_{a_j}\varphi$. Subtracting the holomorphic 1-form $\Sigma\lambda_j\omega_j$ from φ, we can assume that $\int_{a_j}\varphi = 0$ for all j. The preceding formula, with $\omega = \omega_k$, gives

$$(2\pi i)\int_{\gamma}\omega_k = \int_{b_k}\varphi.$$

Write $\gamma = \Sigma m_j a_j + n_j b_j$, with integer coefficients m_j, n_j. Then

$$\begin{aligned}\int_\gamma \omega_k &= \sum_{j=1}^{g}\left(m_j\int_{a_j}\omega_k + n_j\int_{b_j}\omega_k\right)\\ &= m_k + \sum_{j=1}^{g} n_j\int_{b_j}\omega_k = m_k + \sum_{j=1}^{g} n_j\int_{b_k}\omega_j.\end{aligned}$$

the last step by the symmetry of Corollary 20.22. But now if we set $\omega = 2\pi i\Sigma n_j\omega_j$, this shows that $\int_{b_k}(\varphi - \omega) = 2\pi i m_k$. Therefore

$$\int_{a_k}(\varphi - \omega) = -\int_{a_k}\omega = -2\pi i\int_{a_k}\sum n_j\omega_j = -2\pi i n_k,$$

and this completes the proof of the claim, and hence of Abel's theorem.

Theorem 21.32 (Jacobi Inversion). *The Abel–Jacobi map from the group of cycles of degree zero to the Jacobian $J(X)$ is surjective.*

This means that we have an exact sequence

$$0 \to \mathbb{C}^* \to M(X)^* \to \tilde{Z}_0(X) \to J(X) \to 0,$$

which realizes the torus $J(X) = \mathbb{C}^g/\Lambda$ as the quotient of the group of divisors of degree zero by the subgroup of divisors of meromorphic functions.

The proof, which we only sketch, requires a few basic facts about holomorphic mappings of several complex variables. Let X^g be the g-fold Cartesian product of X with itself, which is a g-dimensional complex manifold. Let

$$\alpha_g \colon X^g \to J(X)$$

be the map which takes $(P_1, \ldots, P_g)$ to $A(\Sigma P_i - gP_0)$, where A is the Abel–Jacobi map, and P_0 is any fixed point on X. It suffices to prove that α_g is surjective. The Jacobian $J(X) = \mathbb{C}^g/\Lambda$ gets the structure of a complex manifold so that the quotient mapping $\mathbb{C}^g \to J(X)$ is a local isomorphism.

Exercise 21.33. Verify that α_g is a holomorphic (analytic) mapping of complex manifolds.

We claim next that there are distinct points $P_1, \ldots, P_g$ in X such that the Jacobian determinant of α_g at the point $(P_1, \ldots, P_g)$ is not zero. Once this is verified, it follows that the image of α_g contains an open set. Since the image of the Abel–Jacobi map is a subgroup,

and $J(X)$ is compact, it follows that the image must be all of $J(X)$. To prove the claim, take any distinct points $P_1, \ldots, P_g$. Let z_j be a local coordinate at P_j, and write $\omega_i = f_{i,j}\,dz_j$ near P_j.

Exercise 21.34. Verify that, in suitable coordinates, the Jacobian matrix of α_g at $(P_1, \ldots, P_g)$ is $(f_{i,j}(P_j))$.

Now take the points $P_1, \ldots, P_g$ as in Exercise 21.20(a). If the Jacobian determinant $\det(f_{i,j}(P_j))$ of α_g vanishes at $(P_1, \ldots, P_g)$, then there are g complex numbers $\lambda_1, \ldots, \lambda_g$, not all zero, such that

$$(\lambda_1\omega_1 + \ldots + \lambda_g\omega_g)(P_j) = 0$$

for all j. But this means that $\Omega(P_1 + \ldots + P_g)$ is not zero, which contradicts Exercise 21.20(a).

PART XI

HIGHER DIMENSIONS

This last part is designed to introduce the reader to a few of the higher-dimensional generalizations of the ideas we have studied in earlier chapters, both to unify these ideas, and to indicate a few of the directions one may go if one continues in algebraic topology. It is not written as the culmination or goal of the rest of the course, but rather as a brief introduction to the general theory. How accessible or useful it may be depends on several factors, such as background in manifold theory, and ability to generalize from the special cases we have seen to higher dimensions (an ability, it seems to me, often underestimated in our teaching). For systematic developments of the ideas of this part, the books of Bredon (1993), Bott and Tu (1980), Greenberg and Harper (1981), and Massey (1991) are recommended.

We have studied the first homology group H_1X, the fundamental group $\pi_1(X, x)$, and the first De Rham cohomology group H^1X, which were sufficient to capture most of the topology of the spaces we have been most concerned with: open sets in the plane and surfaces, and an occasional graph. Each of these is the first in a sequence of groups that are used to study similar questions for higher-dimensional spaces. In contrast to many earlier chapters, the tone in this final part is designed to be more formal, concise, and abstract; we are depending on your experience with special cases and low-dimensional examples for motivation.

We start by recalling some three-dimensional calculus, to indicate the sort of "topology" these higher groups might measure. Then we take a quick look at knots in 3-space, mainly because knot theory is

an interesting and important subject in its own right, and also because it gives us a chance to use some of the tools developed in earlier chapters. We also define the higher homotopy groups of a space and the De Rham cohomology groups.

In the next chapter we define higher homology groups, and prove their basic properties. We indicate some of the ways they can be used to extend ideas we have looked at in the plane or on surfaces to higher dimensions. In particular, they give a simple extension of the notion of degree, and they lead to generalizations of the Jordan curve theorem.

In the final chapter we include a couple of "diagram chasing" facts from algebra, one of which can be used to compare different homology and cohomology groups, the other to construct long exact sequences such as Mayer–Vietoris sequences. (Having proved this once and for all, one does not have to keep doing the same sort of manipulations we have done to define boundary and coboundary maps.) Finally, we give proofs of the basic duality theorems between homology and De Rham cohomology on manifolds.

The sections involving higher De Rham cohomology on manifolds are written for those with a working knowledge of differential forms on differentiable manifolds. A reader without this background can skip or skim this part (or stick to low dimensions and/or open sets in $\mathbb{R}^n$). The construction of higher homology groups and its applications to higher-dimensional analogues of the theorems we saw at the beginning of the course do not depend on any of this, however, and a reader who has mastered the earlier chapters should be able to work through this without any gaps. We have included Borsuk's theorem that maps of spheres that preserve antipodal pairs have odd degree, since this allows generalizing all the results we proved about winding numbers to higher dimensions.

A few remarks about other approaches may be in order. It is possible to define the degree of a mapping of a sphere to itself without the notions of homology, and to prove many of its properties. Those with a knowledge of differential topology can do this by approximating a continuous map by a differentiable one, and following the pattern of Problem 3.32 for the winding number. (For a nice discussion in the $\mathscr{C}^\infty$ context, see Milnor (1965)). There are also elementary approaches using simplicial approximations, although considerable care is required to make the arguments rigorous. (In fact, the difficulties in this approach seem to us much greater than those involved in developing general homology theory—not to mention the fact that, having done the latter, one can apply it in many other situations.)

We might also mention that we have used cubes rather than the simplices that are common in many other treatments. Simplices have a slight advantage that one has no "degenerate" maps to ignore, but cubes are simpler for homotopies and product spaces in general, and they are more convenient for integrating differential forms. (In fact, it is not hard—as we do in §24d—to use cubes to calculate the simplicial homology of spaces that are triangulated.)

Finally, a word on terminology: here *manifolds* always are assumed to have a countable basis for their topology.

CHAPTER 22

Toward Higher Dimensions

22a. Holes and Forms in 3-Space

On an open set U in $\mathbb{R}^3$ we have: *0-forms*, which are just $\mathscr{C}^\infty$ functions on U; *1-forms*, which are expressions

$$p\,dx + q\,dy + r\,dz,$$

where p and q and r are $\mathscr{C}^\infty$ functions on U; *2-forms*, which are expressions

$$u\,dy\,dz + v\,dx\,dz + w\,dx\,dy,$$

where u and v and w are $\mathscr{C}^\infty$ functions on U; and *3-forms*, which are expressions

$$h\,dx\,dy\,dz,$$

where h is a $\mathscr{C}^\infty$ function on U.

These forms are designed for integrating, just as in the plane. A 0-form is evaluated at points. The integral of a 1-form over a differentiable path $\gamma\colon [a,b] \to U$ is defined exactly as for the plane:

$$\int_\gamma p\,dx + q\,dy + r\,dz \;=\; \int_a^b \left(p(\gamma(t))\frac{dx}{dt} + q(\gamma(t))\frac{dy}{dt} + r(\gamma(t))\frac{dz}{dt}\right) dt,$$

where $\gamma(t) = (x(t), y(t), z(t))$. A 2-form can be integrated over a dif-

ferentiable map $\Gamma: R \to U$ from a rectangle $R = [a, b] \times [c, d]$ to U:

$$\iint_\Gamma u\,dy\,dz + v\,dx\,dz + w\,dx\,dy =$$

$$\iint_R \left(u(\Gamma(s,t)) \frac{\partial(y,z)}{\partial(s,t)} + v(\Gamma(s,t)) \frac{\partial(x,z)}{\partial(s,t)} + w(\Gamma(s,t)) \frac{\partial(x,y)}{\partial(s,t)} \right) ds\,dt,$$

where, for $\Gamma(s,t) = (x(s,t), y(s,t), z(s,t))$, and $\partial(x,y)/\partial(s,t)$ denotes $\partial x/\partial s\ \partial y/\partial t - \partial x/\partial t\ \partial y/\partial s$, and similarly for the other terms. A 3-form is integrated over a differentiable map $\Pi: B \to U$ where B is a rectangular box $[a, b] \times [c, d] \times [e, f]$:

$$\iiint_\Pi h\,dx\,dy\,dz = \iiint_B h(\Pi(s,t,u)) \frac{\partial(x,y,z)}{\partial(s,t,u)} ds\,dt\,du,$$

where $\Pi(s,t,u) = (x(s,t,u), y(s,t,u), z(s,t,u))$, and $\partial(x,y,z)/\partial(s,t,u)$ denotes the Jacobian determinant.

The *differential* df of a 0-form f is a 1-form defined by

$$df = \frac{\partial f}{\partial x} dx + \frac{\partial f}{\partial y} dy + \frac{\partial f}{\partial z} dz.$$

The fundamental theorem of calculus gives $\int_\gamma df = f(\gamma(b)) - f(\gamma(a))$ for γ a path as above. The differential of a 1-form is a 2-form:

$$d(p\,dx + q\,dy + r\,dz)$$
$$= \left(\frac{\partial r}{\partial y} - \frac{\partial q}{\partial z}\right) dy\,dz + \left(\frac{\partial r}{\partial x} - \frac{\partial p}{\partial z}\right) dx\,dz + \left(\frac{\partial q}{\partial x} - \frac{\partial p}{\partial y}\right) dx\,dy,$$

and Green's theorem (for rectangles in the plane) gives $\iint_\Gamma d\omega = \int_{\partial\Gamma} \omega$, where the integral around the boundary of the rectangle is defined as in Part I. Finally, the differential of a 2-form is a 3-form:

$$d(u\,dy\,dz + v\,dx\,dz + w\,dx\,dy) = \left(\frac{\partial u}{\partial x} - \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z}\right) dx\,dy\,dz,$$

and Stokes theorem says that $\iiint_\Pi d\omega = \iint_{\partial\Pi} \omega$.

Exercise 22.1. State and prove Stokes' theorem for a box, and define the boundary of Π as a sum and difference of the restrictions of Π to the six sides of the box, assigning correct signs to each so the above formula holds.

We define the differential of a 3-form to be 0. A simple calculation

using the equality of mixed derivatives shows that if f is a 0-form, then $d(df) = 0$, and if ω is a 1-form, then $d(d\omega) = 0$. A k-form ω is *closed* if $d\omega = 0$, and *exact* if $\omega = d\mu$ for some $(k-1)$-form μ. So all exact forms are closed, and we have the same question as in the plane: when are closed forms exact? We can define the De Rham groups H^kU as before: for $k = 0, 1, 2$, or 3,

$$H^kU = \{\text{closed } k\text{-forms on } U\}/\{\text{exact } k\text{-forms on } U\}.$$

The question becomes: How does the topology of U influence the size of the vector spaces H^kU?

For 1-forms, the answer is very similar to the case of open sets in the plane. A closed 1-form ω on U is exact if and only if integral of ω over paths in U depend only on the endpoints, or all integrals over closed paths are zero. For example, if U is the complement of the z-axis, the 1-form $\omega = (-y\,dx + x\,dy)/(x^2+y^2)$ is closed but not exact; as we know, an integral of ω around a circle in the xy-plane is 2π. Notice that taking a point or a closed ball out of $\mathbb{R}^3$ does not count as a "hole" as far as 1-forms in 3-space is concerned. In fact:

Exercise 22.2. Show that if $H_1U = 0$ then every closed 1-form on U is exact.

For 2-forms, however, if U is the complement of a point, there are closed forms that are not exact. For example, let

$$\omega = \frac{x\,dy\,dz - y\,dx\,dz + z\,dx\,dy}{(x^2+y^2+z^2)^{3/2}}.$$

Exercise 22.3. Show that ω is closed. Fix a positive number ρ, and let Γ: $[0, 2\pi] \times [-{}^1\!/_2\pi, {}^1\!/_2\pi] \to \mathbb{R}^3$ be the spherical coordinate mapping, i.e.,

$$\Gamma(\vartheta, \varphi) = (\rho\cos(\vartheta)\cos(\varphi), \rho\sin(\vartheta)\cos(\varphi), \rho\sin(\varphi)).$$

Compute the integral of ω over Γ, and deduce that ω is not exact.

Exercise 22.4. Use this 2-form ω to define the *engulfing number* around 0 of a differentiable map from S^2 to $\mathbb{R}^3 \setminus \{0\}$. Can you prove any analogues of the winding number?

We will see in the next chapter how to define second homology groups H_2U that have the same relation to 2-forms and H^2U as the first homology H_1U has to 1-forms and H^1U.

Some of these ideas may be more familiar in vector field language:

a 1-form $p\,dx + q\,dy + r\,dz$ can be identified with the vector field $p\mathbf{i} + q\mathbf{j} + r\mathbf{k}$, the 2-form $u\,dy\,dz + v\,dx\,dz + w\,dx\,dy$ with the vector field $u\mathbf{i} - v\mathbf{j} + w\mathbf{k}$, and the 3-form $h\,dx\,dy\,dz$ with the function h. In this language, the differential df of a function corresponds to the *gradient* grad(f), the differential of a 1-form to the *curl* of a vector field, and the differential of a 2-form becomes the *divergence* of a vector field:

$$\begin{aligned}
\operatorname{grad}(f) &= \frac{\partial x}{\partial x}\mathbf{i} + \frac{\partial f}{\partial y}\mathbf{j} + \frac{\partial f}{\partial z}\mathbf{k};\\
\operatorname{curl}(p\mathbf{i} + q\mathbf{j} + r\mathbf{k}) &= \left(\frac{\partial r}{\partial y} - \frac{\partial q}{\partial z}\right)\mathbf{i} + \left(\frac{\partial p}{\partial z} - \frac{\partial r}{\partial x}\right)\mathbf{j} + \left(\frac{\partial q}{\partial x} - \frac{\partial p}{\partial y}\right)\mathbf{k};\\
\operatorname{div}(u\mathbf{i} + v\mathbf{j} + w\mathbf{k}) &= \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z};
\end{aligned}$$

and the equations $d \circ d = 0$ say that $\operatorname{curl} \circ \operatorname{grad} = 0$ and $\operatorname{div} \circ \operatorname{curl} = 0$. The integral of the vector field corresponding to a 2-form over a surface can be interpreted as the integral of a dot product with an outward-pointing normal.

Problem 22.5. (a) Generalize the discussion of §2c to fluid flows on open sets in 3-space. Interpret the engulfing number as the flux across a surface of a flow with source at the origin (see Exercise 2.26). (b) Define harmonic functions of three variables, and generalize Exercise 2.21 and Problems 2.22–2.25.

22b. Knots

A *knot* is a subset K of $\mathbb{R}^3$ or the 3-sphere S^3 that is homeomorphic to a circle. Call two knots *equivalent* if there is a homeomorphism of $\mathbb{R}^3$ (or S^3) with itself that takes one homeomorphically onto the other, and is orientation-preserving. (For a precise definition of "orientation-preserving," see §23c.) A weaker notion is *similarity*, which is the same except ignoring orientation, so "mirror image" knots are always similar. The generalized Jordan curve theorem, which we will prove in the next chapter, implies that the complement of X is connected, that its first homology group is infinite cyclic, and its other homology groups vanish. In particular, these groups are the same for all knots. However, the fundamental group $\pi_1(S^3 \setminus X)$ or $\pi_1(\mathbb{R}^3 \setminus X)$ can sometimes be used to distinguish knots from each other. Note that these fundamental groups are the same for similar knots, so they can be used as a possible invariant.

Exercise 22.6. If $K \subset \mathbb{R}^3$ is a knot, and $\mathbb{R}^3$ is identified with the complement of a point in S^3, by stereographic projection, show that the fundamental group of $\mathbb{R}^3 \setminus K$ is isomorphic to the fundamental group of $S^3 \setminus K$.

The standard embedding $K = S^1 \subset \mathbb{R}^2 \subset \mathbb{R}^3 \subset S^3$ is called the trivial knot, and any knot equivalent to this is called *trivial*. We want to use fundamental groups to give one example of a knot that is not trivial. We will do this by identifying $S^3 \subset \mathbb{R}^4 = \mathbb{C}^2$ with the set of pairs (z, w) of complex numbers such that $|z|^2 + |w|^2 = 1$, and constructing a knot K that is the intersection of S^3 with a complex plane curve that has a singularity at the origin. This is more than a convenient way to find an example: the knots that arise this way are important invariants of singularities of plane curves! In this language, the trivial knot can be realized as the intersection of S^3, with the "curve" $w = 0$.

Exercise 22.7. Show that the fundamental group of the complement of the trivial knot is isomorphic to $\mathbb{Z}$. Show, in fact, that the circle $\{(0, w): |w| = 1\}$ is a deformation retract of the complement of circle $\{(z, 0): |z| = 1\}$ in S^3.

In nature coverings often arise by starting with a mapping that is not a covering, but becomes one after throwing away a locus where it fails to be a covering. We have seen this for an analytic mapping between Riemann surfaces in Chapter 19, where only a finite set had to be thrown away. With appropriate hypotheses, such a mapping is called a "branched covering," with the bad set the "branch locus." Here is another example. Consider the mapping $f\colon \mathbb{C}^2 \to \mathbb{C}^2$ given by the formula $f(u, v) = (u, v^3 + uv)$. The inverse image of a point (z, w) has three points if the equation $v^3 + zv = w$ has three distinct solutions for the variable v, and one or two points otherwise.

Exercise 22.8. Show that $f^{-1}(z, w)$ has three points if and only if $4z^3 + 27w^2 \neq 0$. If $4z^3 + 27w^2 = 0$, but $(z, w) \neq (0, 0)$, the inverse image has two points, and for $(z, w) = (0, 0)$, the inverse image has one point.

Let $V \subset \mathbb{C}^2$ be the plane curve $4z^3 + 27w^2 = 0$, which is the branch locus of the above mapping f. Let K be the intersection of V with S^3:

$$V = \{(z, w): 4z^3 + 27w^2 = 0\}, \qquad K = V \cap S^3.$$

We claim first that K is homeomorphic to a circle.

Exercise 22.9. Show that the mapping $e^{2\pi it} \mapsto (-ae^{4\pi it}, be^{6\pi it})$, where a is the positive solution to the equation $4a^3 + 27a^2 = 27$ and $b = \sqrt{1 - a^2}$, is a homeomorphism of S^1 onto K.

We will consider the mapping $\mathbb{C}^2 \setminus f^{-1}(V) \to \mathbb{C}^2 \setminus V$ determined by f, and the restriction

$$p\colon Y = f^{-1}(S^3 \setminus K) \to S^3 \setminus K = X, \qquad (u, v) \mapsto (u, v^3 + uv).$$

Take $x = (1, 0)$ as the base point in X, and $y = (1, 0)$ as the base point in Y. Note that $p^{-1}(x) = \{(1,0), (1,i), (1,-i)\}$.

Claim 22.10. (1) *p is a three-sheeted covering map*; (2) *Y is connected*; *and* (3) *p is not a regular covering.*

It follows from this claim that $\pi_1(S^3 \setminus K, x)$ is not abelian, since every connected covering of a manifold with an abelian fundamental group is regular. In particular, K is not a trivial knot. We leave the proofs of (1)–(3) as exercises, with the following comments. The essential point of (1) is showing that the roots of a polynomial are, locally where the roots are distinct, continuous functions of the coefficients. In fact, they are analytic functions, by the same argument as in §20a. For (2), it suffices to show that the three points of $f^{-1}(x)$ can be connected by paths. Consider the loop $\gamma(t) = (e^{2\pi it}, 0)$, $0 \le t \le 1$, at x. This lifts to the loop $\tilde{\gamma}_1(t) = (e^{2\pi it}, 0)$ at y, and to the path $\tilde{\gamma}_2(t) = (e^{2\pi it}, ie^{\pi it})$ that goes from $(1, i)$ to $(1, -i)$ in Y.

Exercise 22.11. Find a path of the form $\sigma(t) = (\lambda(t), i\mu(t))$, $0 \le t \le 1$, with $\lambda(t) > 0$ and $\mu(t) \ge 0$ for all t, that goes from $(1, 0)$ to $(1, i)$ in Y.

This exercise shows that Y is connected, and since $\tilde{\gamma}_1$ is closed, and $\tilde{\gamma}_2$ is not, the covering is not regular.

In fact, up to equivalence, K is an example of a *torus knot*. The torus $T = \mathbb{R}^2/\mathbb{Z}^2$ sits in S^3 by the mapping that takes (x, y) to $((1/\sqrt{2})e^{2\pi ix}, (1/\sqrt{2})e^{2\pi iy})$. For any relatively prime pair of positive integers p and q, the image in the torus of the line with equation $qy = px$ in $\mathbb{R}^2$ is a knot that winds p times around the torus one way while it winds q times around the other way. This is called a torus knot of type (p, q).

Exercise 22.12. (a) Show that the knot $K = V \cap S^3$ considered above is equivalent to a torus knot of type $(2, 3)$. (b) Show that for relatively prime positive integers p and q the intersection of $z^q = w^p$ with S^3 is a torus knot of type (p, q).

The Van Kampen theorem can be used to calculate the fundamental group of the complement of any torus knot. The 3-sphere is the union of two solid tori

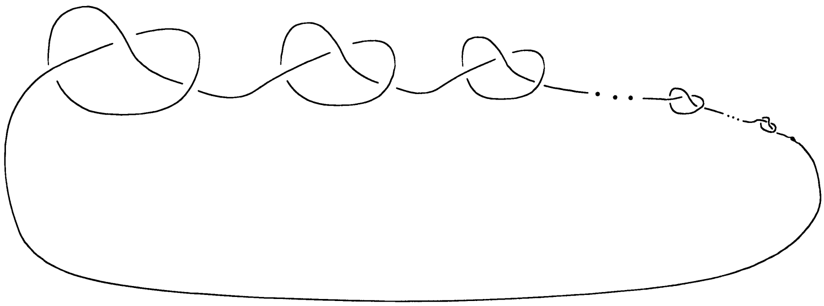
$$A = \{(z, w) \in S^3 \colon |z| \leq |w|\}, \qquad B = \{(z, w) \in S^3 \colon |z| \geq |w|\},$$

so $T = A \cap B = \{(z, w) \colon |z| = |w| = 1/\sqrt{2}\}$ is a torus.

Problem 22.13. If K is a torus knot of type (p, q) in T, show that the fundamental groups of $A \setminus K$, $B \setminus K$, and $T \setminus K$ are infinite cyclic, and the generator of the fundamental group of $T \setminus K$ maps to the pth and qth powers of generators of the fundamental groups of $A \setminus K$ and $B \setminus K$. Apply the Van Kampen theorem to show that the fundamental group of $S^3 \setminus K$ has two generators a and b, and one relation $a^p \cdot b^q = e$, i.e., the fundamental group is F_2/N, where F_2 is the free group on a and b, and N is the least normal subgroup containing $a^p \cdot b^q$.

For a knot of type $(2, 3)$, for example, one can see again that this group is not abelian by mapping it onto the symmetric group $\mathfrak{S}_3$ on three letters, sending a to the transposition $(1\ 2)$ and b to the permutation $(1\ 2\ 3)$.

There are many knots that are not torus knots. For example, one can take a torus knot, and take a small tube around it, which is homeomorphic to another torus, and put a torus knot on this. Repeating this construction arbitrarily often gives a class of knots which, remarkably, are exactly the knots one gets from singularities of plane curves. There are many other knots, however. Moreover, there are some "wild" knots, such as "Antoine's necklace":



The fundamental group of the complement of this knot is not even finitely generated. A piece of this is an embedding of a closed interval in $\mathbb{R}^3$ such that the complement is not simply connected.

22c. Higher Homotopy Groups

The higher homotopy groups $\pi_k(X, x)$ are easier to define than higher homology or cohomology groups, although their calculation turns out to be far more challenging. Fix a base point s_0 in the sphere S^k, say the north pole: $s_0 = (0, \ldots, 0, 1)$. Define $\pi_k(X, x)$ to be the set of homotopy classes of maps from S^k to X that map s_0 to x; here a homotopy between two such maps must preserve basepoints throughout the homotopy, i.e., H is a continuous map from $S^k \times [0, 1]$ to X, with $H(s_0 \times t) = x$ for all $0 \le t \le 1$. One can also define $\pi_k(X, x)$ as the set of homotopy classes of maps from the standard k-cube I^k to X that map the boundary of the cube to x, with homotopies also mapping the boundary to x throughout.

Exercise 22.14. (a) Show that these two definitions agree by showing that S^k is homeomorphic to the space obtained by identifying all points of the boundary of I^k to a point. (b) Show that $\pi_k(X, x) = 0$ for all $k > 0$ if X is contractible. (c) Show that a map $f: X \to Y$ determines maps $f_*: \pi_k(X, x) \to \pi_k(Y, f(x))$, which are functorial, and that maps that are homotopic through basepoint-preserving homotopies determine the same map on homotopy groups. (d) Show that $\pi_k(S^n, s_0) = 0$ for $0 < k < n$.

The sets $\pi_k(X, x)$ can be made into groups, much as for the fundamental group. Using the definition by cubes, one can "multiply" two maps Γ and Λ from I^k to X, defining $\Gamma \cdot \Lambda$ by using the first coordinate:

$$\Gamma \cdot \Lambda(t_1, \ldots, t_k) = \begin{cases} \Gamma(2t_1, t_2, \ldots, t_k), & 0 \le t_1 \le 1/2, \\ \Lambda(2t_1 - 1, t_2, \ldots, t_k), & 1/2 \le t_1 \le 1. \end{cases}$$

Exercise 22.15. (a) Show that this operation is well defined on homotopy classes, and makes $\pi_k(X, x)$ into a group. (b) Show that the maps f_* of Exercise 22.14 are homomorphisms of groups.

Problem 22.16. Show that, for all $k > 1$, the group $\pi_k(X, x)$ is abelian.

It is a fact that $\pi_n(S^n, s_0) \cong \mathbb{Z}$, although this is quite a bit harder to prove. Note that this gives a strong notion of degree for maps of S^n to S^n: it defines the degree, and shows that maps are classified up to homotopy by their degree. In the next chapter we will use chains to

define homology groups $H_k(X)$, which are easier to calculate, and we will show that $H_n(S^n) = \mathbb{Z}$ and $H_k(S^n) = 0$ for all $k > 0$, $k \neq n$. In stark contrast with the homology groups, for $k > n$, the groups $\pi_k(S^n, s_0)$ need not be trivial.

Exercise 22.17. Identifying S^3 with $\{(z, w) \in \mathbb{C}^2: |z|^2 + |w|^2 = 1\}$, show that the map that takes (z, w) to $w/z \in \mathbb{C} \subset S^2$ determines a continuous mapping from S^3 to S^2.

It is a fact that $\pi_3(S^2, s_0) \cong \mathbb{Z}$, with generator given by the mapping of the preceding exercise, which is called a *Hopf* mapping (see Hilton (1961)). We will show in the next chapter that the fibers of the Hopf mapping are circles that are "linked" together in S^3, which is at least an indication of the nontriviality of the Hopf map. It should be pointed out, however, that in spite of enormous effort, which have produced calculations of many special cases, the groups $\pi_k(S^n, s_0)$ are far from known in general.

22d. Higher De Rham Cohomology

All of the discussion of §22a generalizes to n variables, a k-form being an expression

$$\Sigma f_{i_1 i_2 \ldots i_k} dx_{i_1} dx_{i_2} \ldots dx_{i_k} = \Sigma f_I dx_I$$

the sum over all $1 \leq i_1 < \ldots < i_k \leq n$, with the coefficients $f_{i_1 i_2 \ldots i_k}$ $\mathscr{C}^\infty$ functions on an open set in $\mathbb{R}^n$. There is a differential d that takes a k-form to a $(k+1)$-form, and again $d \circ d = 0$. For U open in $\mathbb{R}^n$ one then gets De Rham groups $H^k U$, the space of closed k-forms modulo the space of exact k-forms. The main complication is in keeping track of the signs. This is best done by introducing formally the "exterior algebra" structure that is already apparent in the plane and 3-space: one allows the differentials to be written in arbitrary order, but put in a sign whenever they are interchanged: $dy\,dx = -dx\,dy$, together setting $dx\,dx = 0$. (The usual notation for this exterior product is the wedge "$\wedge$," so one writes $dx \wedge dy$ in place of our $dx\,dy$.) Working this out properly belongs in an advanced calculus course.

As with surfaces, one can define k-forms on an arbitrary differentiable manifold as collections of k-forms on the coordinate neighborhoods of a chart that transform properly on overlaps. If you know about differential forms on a manifold, it is not difficult to generalize

the idea of De Rham cohomology. Again there are differentials d from k-forms to $(k+1)$-forms, with $d \circ d = 0$. In this section we will assume familiarity with notions of manifolds and differential forms. If this applies to you, fine; if not, you can either stick to low dimensions where we have done it by hand, or assume this formalism in general—or you can turn immediately to the next chapter, which does not depend on any of this.

If X is a differentiable manifold, one can define the De Rham cohomology group H^kX as the vector space of closed k-forms modulo the subspace of exact k-forms. If $f\colon X \to Y$ is a differentiable map, there is a notion of pull-back of k-forms from Y to X, taking ω on Y to $f^*\omega$ on X. This commutes with the differential, so determines a (functorial) homomorphism $f^*\colon H^kY \to H^kX$.

Exercise 22.18. If X is a disjoint union of a finite or infinite number of manifolds X_i, show that H^kX is the direct product of the H^kX_i, i.e., specifying a class on X is the same as specifying a class on each X_i.

The Mayer–Vietoris exact sequence is defined with almost no change from the case with H^0 and H^1. To define

$$\delta\colon H^k(U \cap V) \to H^{k+1}(U \cup V),$$

as before, one uses a partition of unity subordinate to an open covering $X = U \cup V$ by U and V to write a closed k-form ω on $U \cap V$ as the difference $\mu_1 - \mu_2$ of a k-form μ_1 on U and a k-form μ_2 on V; there is a closed $(k+1)$-form on $U \cup V$ that is $d\mu_1$ on U and $d\mu_2$ on V, and this $(k+1)$-form represents the image in $H^{k+1}(U \cup V)$ of the class represented by ω in $H^k(U \cap V)$. As before (and see §24a), one proves:

Mayer–Vietoris Theorem 22.19. *For any open sets U and V in a manifold of dimension n, there is an exact sequence*

$$\begin{aligned}
0 \longrightarrow H^0(U \cup V) &\xrightarrow{+} H^0U \oplus H^0V \xrightarrow{-} H^0(U \cap V) \\
\xrightarrow{\delta} H^1(U \cup V) &\xrightarrow{+} H^1U \oplus H^1V \xrightarrow{-} H^1(U \cap V) \xrightarrow{\delta} \\
&\cdots \longrightarrow \cdots \\
\xrightarrow{\delta} H^n(U \cup V) &\xrightarrow{+} H^nU \oplus H^nV \xrightarrow{-} H^n(U \cap V) \longrightarrow 0
\end{aligned}$$

To calculate these groups, one needs in addition the

Poincaré Lemma 22.20. *If $p\colon X \times \mathbb{R} \to X$ is the projection, then $p^*\colon H^kX \to H^k(X \times \mathbb{R})$ is an isomorphism.*

The inverse isomorphism to p^* is $s^*: H^k(X \times \mathbb{R}) \to H^k(X)$, where s is the embedding $x \mapsto x \times 0$ of X in $X \times \mathbb{R}$. The problem is to show that $p^* \circ s^* = (s \circ p)^*$ is the identity. The idea of the proof is to construct a linear map H from the space of k-forms on $X \times \mathbb{R}$ to the space of $(k-1)$-forms on $X \times \mathbb{R}$, for each k, such that for any form ω on $X \times \mathbb{R}$,

$$\omega - p^* \circ s^*(\omega) = d(H(\omega)) + H(d(\omega)). \tag{22.21}$$

(Note that the two H's and the two d's in this equation are defined on different spaces of forms!) It follows that if ω is closed, then $\omega - p^* \circ s^*(\omega) = d(H(\omega))$, so ω and $p^* \circ s^*(\omega)$ define the same De Rham cohomology class.

Problem 22.22. Show that any k-form ω on $X \times \mathbb{R}$ has a unique expression as a sum of a k-form not involving dt, where t is the coordinate on $\mathbb{R}$, and one of the form $dt \wedge \mu$, where μ is, in local coordinates, a sum of expressions $f \cdot dx_I$, with the x_i coordinates on X, and f is a function on the product of the coordinate neighborhood with $\mathbb{R}$. Define H of such a form to be the form obtained by integrating μ with respect to the variable in $\mathbb{R}$ (so forms not involving dt are mapped to 0). For example, if $X \subset \mathbb{R}^n$, and μ is the form $f dx_I$, then $H(dt \wedge \mu)$ is the form $g\, dx_I$, where

$$g(x_1, \ldots, x_n, t) = \int_0^t f(x_1, \ldots, x_n, s)\, ds.$$

Show that this operator is well defined and satisfies (22.21).

For any real number t, if $s_t: X \to X \times \mathbb{R}$ maps x to $x \times t$, then, since $p \circ s_t$ is the identity, it follows that $s_t^*: H^k(X \times \mathbb{R}) \to H^k(X)$ is the inverse to $p^*: H^k(X) \to H^k(X \times \mathbb{R})$. In particular, the maps s_t^* are the same for all t. This implies that if $F: X \times \mathbb{R} \to Y$ is differentiable, all the maps $F_t: X \to Y$, $F_t(x) = F(x \times t)$, determine the same maps $F_t^*: H^k Y \to H^k X$. Indeed, $F_t = F \circ s_t$, so $F_t^* = s_t^* \circ F^*$ is independent of t.

Problem 22.23. (a) Use the Poincaré lemma and Mayer–Vietoris to calculate the De Rham cohomology of $\mathbb{R}^n$, S^n, and $\mathbb{R}^n \setminus \{0\}$. (b) Show that the $(n-1)$-form ω_{n-1} defined on $\mathbb{R}^n \setminus \{0\}$ by

$$\omega_{n-1} = \frac{\sum_{i=1}^{n} (-1)^{i-1} x_i\, dx_1 \wedge \ldots \wedge dx_{i-1} \wedge dx_{i+1} \wedge \ldots \wedge dx_n}{(x_1^2 + \ldots + x_n^2)^{n/2}}$$

is closed and gives a generator of $H^{n-1}(\mathbb{R}^n \setminus \{0\})$. If $f: S^{n-1} \to \mathbb{R}^n \setminus \{0\}$ is differentiable, this can be used to define a higher-dimensional winding number, or "engulfing number": $W(f, 0)$ is the integral of $f^*\omega_{n-1}$ over S^{n-1}, divided by the integral of ω_{n-1} over S^{n-1}.

22e. Cohomology with Compact Supports

There is another way to use differential forms to construct cohomology groups, for open sets in $\mathbb{R}^n$ or any $\mathscr{C}^\infty$ manifolds, which we sketch briefly here. These cohomology groups are called *De Rham groups with compact supports*, and denoted $H_c^k X$, the subscript c standing for "compact." These are defined exactly as for the ordinary De Rham groups, but using differential forms with compact support, i.e., forms for which there is some compact set K contained in X such that the form vanishes identically outside K. Define $H_c^k X$ to be the quotient space of the closed k-forms with compact support by the subspace of forms that are differentials of $(k-1)$-forms with compact support.

If X is compact, of course, all forms have compact support, so $H_c^k X = H^k X$. In spite of this and the similarity of definition, however, these groups are quite different on noncompact manifolds. For example, locally constant functions on a noncompact connected space can never have compact support unless they are identically zero:

$$H_c^0 X = 0 \quad \text{if } X \text{ is connected and not compact}.$$

In fact, we will see that the theories H^k and H_c^k behave in an opposite, or dual, way. By using a partition of unity as in Chapter 18, if X is an oriented n-manifold, one can integrate an n-form with compact support over the whole manifold. Using Stokes' theorem, it follows similarly that the integral of the differential of a closed $(n-1)$-form with compact support is zero, so one has a map

$$H_c^n X \to \mathbb{R}, \qquad \omega \mapsto \int_X \omega .$$

It is easy to produce n-forms that are positive on a small piece of a coordinate neighborhood, and zero elsewhere, to see that this map is not zero. We will see in Chapter 24 that, if X is connected as well, then this map is an isomorphism.

Exercise 22.24. If X is a disjoint union of a finite or infinite number of open sets X_i, show that $H_c^k X$ is the direct sum of the $H_c^k X_i$, i.e.,

specifying a class on X is the same as specifying a class on each X_i, except that all but a finite number must be zero.

There is a Mayer–Vietoris exact sequence for cohomology with compact supports, but it is different from that without supports. First of all, there are no restriction maps, since if ω has compact support on an open set, its restriction to an open subset may no longer have compact support. In fact, the maps go the other way: if U_1 is an open subset of U_2, any k-form ω with compact support on U_1 extends by zero outside U_1 to define a k-form with (the same) compact support on U_2. Since any point in $U_2 \setminus U_1$ has a neighborhood not meeting the support of ω, this extension is $\mathscr{C}^\infty$. This extension commutes with the differential map d, so determines a linear map

$$H^k_c(U_1) \to H^k_c(U_2).$$

In particular, for two open sets U and V, we have diagrams

$$\begin{array}{ccccc} & & H^k_c(U) & & \\ & \nearrow & & \searrow & \\ H^k_c(U \cap V) & & & & H^k_c(U \cup V)\,. \\ & \searrow & & \nearrow & \\ & & H^k_c(V) & & \end{array}$$

We want to define a coboundary map $\delta\colon H^k_c(U \cup V) \to H^{k+1}_c(U \cap V)$. Given a class in $H^k_c(U \cap V)$, represent it by a closed k-form ω with compact supports. We can write ω as a sum $\omega_1 + \omega_2$, with ω_1 and ω_2 k-forms with compact supports in U and V, respectively. In fact, if $\psi_1 + \psi_2 = 1$ is a partition of unity subordinate to U and V, then $\omega_1 = \psi_1 \cdot \omega$ and $\omega_2 = \psi_2 \cdot \omega$ are such forms. Let η be the $(k+1)$-form on $U \cap V$ that is (the restriction of) $d\omega_1$. From the equation $0 = d\omega = d\omega_1 + d\omega_2$, we have $d\omega_1 = -d\omega_2$ on $U \cap V$, so the support of η is contained in the intersection of the supports of ω_1 and ω_2. In particular, η has compact support on $U \cap V$. Clearly η is closed. Set

$$\delta([\omega]) = [\eta].$$

It is not hard to verify that this is independent of the choice, and it is the familiar (by now) argument (and see §24a) to prove the

Mayer–Vietoris Theorem 22.25 (Compact Supports). *For open sets U and V in an n-dimensional manifold, there is an exact sequence*

$$0 \longrightarrow H_c^0(U \cap V) \xrightarrow{+} H_c^0 U \oplus H_c^0 V \xrightarrow{-} H_c^0(U \cup V)$$
$$\xrightarrow{\delta} H_c^1(U \cap V) \xrightarrow{+} H_c^1 U \oplus H_c^1 V \xrightarrow{-} H_c^1(U \cup V) \xrightarrow{\delta}$$
$$\cdots \longrightarrow \cdots$$
$$\xrightarrow{\delta} H_c^n(U \cap V) \xrightarrow{+} H_c^n U \oplus H_c^n V \xrightarrow{-} H_c^n(U \cup V) \longrightarrow 0.$$

As before, to complete the basic tools for calculating these groups, we need to compare a manifold X and $X \times \mathbb{R}$. This time the projection p from $X \times \mathbb{R}$ to X determines homomorphisms

$$p_*\colon H_c^k(X \times \mathbb{R}) \to H_c^{k-1}(X),$$

by "integrating along the fiber," as follows. As in Problem 22.22, one can write a k-form with compact support of $X \times \mathbb{R}$ as a sum of a form not involving dt, and a k-form $dt \wedge \mu$, where μ is, in local coordinates, a sum of expressions $f \cdot dx_I$, with the x_i coordinates on X, and f a function on $X \times \mathbb{R}$. Define p_* of such a form $dt \wedge f \cdot dx_I$ to be $g \cdot dx_I$, where

$$g(x_1, \ldots, x_n) = \int_{-\infty}^{\infty} f(x_1, \ldots, x_n, t)\, dt$$

(so forms not involving dt are mapped to zero). Note that these integrals are really over finite intervals, by the assumption of compact support. One checks that this is well defined, and that $p_*(d\omega) = d(p_*\omega)$, so p_* determines a map on cohomology classes as indicated. If $s\colon X \to X \times \mathbb{R}$ is the inclusion $x \mapsto x \times 0$, there is a map

$$s_*\colon H_c^{k-1}(X) \to H_c^k(X \times \mathbb{R})$$

determined by sending a form ω to $\rho(t)\, dt \wedge \omega$, where ρ is any function with compact support on $\mathbb{R}$ such that $\int_{-\infty}^{\infty} \rho(t)\, dt = 1$. One checks that this commutes with d, so defines a map on cohomology, and that $p_* \circ s_*$ is the identity.

Poincaré Lemma 22.26 (Compact Supports). *For any manifold X, $p_*\colon H_c^k(X \times \mathbb{R}) \to H_c^{k-1}(X)$ is an isomorphism.*

Problem 22.27. Prove this by constructing an operator H from k-forms with compact support to $(k-1)$-forms with compact support on $X \times \mathbb{R}$. This operator should vanish on forms without "dt," and take $dt \wedge \mu$, where μ is, in local coordinates, a sum of expressions

$f \cdot dx_I$, to $g \cdot \mu$, with

$$g(x_1, \ldots, x_n, t) = \int_{-\infty}^{t} f(x_1, \ldots, x_n, s)\, ds$$

$$- \int_{-\infty}^{t} \rho(s)\, ds \int_{-\infty}^{\infty} f(x_1, \ldots, x_n, s)\, ds.$$

Show that for any k-form ω with compact support, $\omega - s_* p_* \omega = d(H(\omega)) + H(d(\omega))$, and deduce that p_* and s_* determine inverse isomorphisms.

Exercise 22.28. Calculate $H_c^k X$, when X is $\mathbb{R}^n$, S^n, and $\mathbb{R}^n \setminus \{0\}$.

CHAPTER 23

Higher Homology

23a. Homology Groups

The groups H_0X and H_1X are the beginning of a series of abelian groups H_kX, defined for any topological space X. Define a *k-cube* in X to be a continuous map $\Gamma: I^k \to X$, where I^k is the k-dimensional cube, i.e., $I = [0, 1]$, so

$$I^k = [0,1] \times \ldots \times [0,1] \subset \mathbb{R}^k.$$

For any such mapping Γ, and any integer i between 1 and k, and any $0 \leq s \leq 1$, define a $(k-1)$-cube $\partial_i^s\Gamma$, which is obtained by restricting Γ to the slice of the ith coordinate at s:

$$\partial_i^s\Gamma: I^{k-1} \to X, \quad \partial_i^s\Gamma(t_1, \ldots, t_{k-1}) = \Gamma(t_1, \ldots, t_{i-1}, s, t_i, \ldots, t_{k-1}).$$

Call Γ *degenerate* if, for some i, $\partial_i^s\Gamma$ is a constant function of s, and *nondegenerate* otherwise. (When $k = 1$, Γ is a path, and degenerate is the same as a constant path.) By convention, $I^0 = \{0\}$, so as 0-cube is given by a point in X; no 0-cube is regarded as degenerate.

Let C_kX be the free abelian group on the nondegenerate k-cubes in X, so an element of C_kX is a finite linear combination $\Sigma n_\alpha \Gamma_\alpha$, with Γ_α a k-cube in X and n_α an integer. An element of C_kX is called a cubical *k-chain* on X. It is useful to regard any finite linear combination $\Sigma n_\alpha \Gamma_\alpha$ of arbitrary k-cubes as an element of C_kX, by simply discarding any degenerate k-cube Γ_α that appears. (In other words, C_kX is identified with the quotient of the free abelian group on all k-cubes in X, modulo the subgroup generated by degenerate k-cubes.)

If $\Gamma: I^k \to X$ is a k-cube in X, its *boundary* $\partial\Gamma$ in $C_{k-1}X$ is defined by the formula

$$\partial\Gamma = \sum_{i=1}^{k} (-1)^i(\partial_i^0\Gamma - \partial_i^1\Gamma),$$

which is the sum of the $2k$ faces of Γ, each with a coefficient of $+1$ or -1. (Note that, even if Γ is nondegenerate, some of the $\partial_i^0\Gamma$ or $\partial_i^1\Gamma$ occurring in the formula can be degenerate, so they are discarded.) This is extended linearly to a homomorphism

$$\partial: C_kX \to C_{k-1}X$$

by the formula $\partial(\Sigma n_\alpha\Gamma_\alpha) = \Sigma n_\alpha(\partial\Gamma_\alpha)$. A k-chain is called a *k-cycle* if its boundary is zero; the k-cycles form a subgroup Z_kX of C_kX. The boundaries of $(k+1)$-chains form a subgroup B_kX of C_kX.

Exercise 23.1. Show that for any $(k+1)$-cube Γ, $\partial(\partial\Gamma) = 0$ in $C_{k-1}X$.

From this exercise it follows that B_kX is a subgroup of Z_kX, so we can define the kth homology group of X to be quotient

$$H_kX = Z_kX/B_kX.$$

Exercise 23.2. (a) Show that $H_kX = 0$ if X is a point and $k > 0$. (Note that this would not be true if degenerate cubes had not been discarded.) (b) Verify that for $k = 0$ and $k = 1$, these are the groups we studied in Chapter 6. (c) Show how any continuous mapping $f: X \to Y$ determines homomorphisms $f_*: H_kX \to H_kY$, and show that these are functorial in the sense of Exercise 6.20. (d) Construct homomorphisms from $\pi_k(X, x)$ to H_kX that are compatible with the maps of (c) and Exercise 22.14.

Proposition 23.3. *If f and g are homotopic maps from X to Y, then f_* and g_* determine the same homomorphisms from H_kX to H_kY.*

Proof. Suppose $H: X \times [0,1] \to Y$ is a homotopy from f to g, and $\Gamma: I^k \to X$ is a k-cube; define a $(k+1)$-cube $R(\Gamma)$ by the formula

$$R(\Gamma)(s, t_1, \ldots, t_k) = H(\Gamma(t_1, \ldots, t_k) \times s).$$

If $\Sigma n_\alpha\Gamma_\alpha$ is a k-cycle, a little calculation shows that the boundary of $\Sigma n_\alpha R(\Gamma_\alpha)$ is $\Sigma n_\alpha(f \circ \Gamma_\alpha) - \Sigma n_\alpha(g \circ \Gamma_\alpha)$, which completes the proof. A more elegant way to see this is to extend R by linearity to a map $R: C_kX \to C_{k+1}X$, $R(\Sigma n_\alpha\Gamma_\alpha) = \Sigma n_\alpha R(\Gamma_\alpha)$. Then a formal calculation shows

that

$$g_* - f_* = \partial \circ R + R \circ \partial$$

as homomorphisms from C_kX to C_kX. Then if z is a k-cycle,

$$g_*(z) - f_*(z) = \partial \circ R(z) + R \circ \partial(z) = \partial(R(z)),$$

which shows that $g_*(z)$ and $f_*(z)$ differ by a boundary. □

It follows from the proposition that if $Y \subset X$ is a deformation retract, then $H_kY \to H_kX$ is an isomorphism for all k. For example, if X is contractible, then $H_kX = 0$ for all $k > 0$.

23b. Mayer–Vietoris for Homology

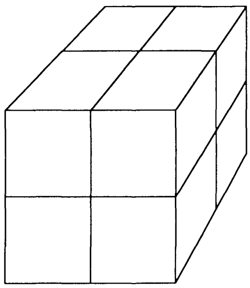
To calculate the higher homology groups of more interesting spaces, we want to extend the Mayer–Vietoris sequence to these higher groups: if U and V are open subsets of a space, there is an exact sequence

$$\cdots \xrightarrow{\partial} H_k(U \cap V) \xrightarrow{-} H_kU \oplus H_0V \xrightarrow{+} H_k(U \cup V)$$

$$\xrightarrow{\partial} H_{k-1}(U \cap V) \xrightarrow{-} H_{k-1}U \oplus H_{k-1}V \xrightarrow{+} H_{k-1}(U \cup V) \xrightarrow{\partial} \cdots .$$

The idea is very much as in Chapter 10. The key is the definition of the "boundary maps" $\partial\colon H_k(U \cup V) \to H_{k-1}(U \cap V)$. For this, we show that any class in $H_k(U \cup V)$ can be represented by a k-cycle z that is a sum of a chain c_1 on U and a chain c_2 on V. Then $\partial c_1 = -\partial c_2$ is a cycle on $U \cap V$, and this represents the image class in $H_{k-1}(U \cap V)$.

To carry out the construction of the boundary map, we need a systematic way to subdivide k-cubes and chains into sums of smaller chains, generalizing the constructions we used in Chapter 10. Given a k-cube $\Gamma\colon I^k \to X$, we define $S(\Gamma)$ to be the sum of the 2^k k-cubes obtained by restricting Γ to each of the 2^k subcubes obtained by subdividing the cube:



Each of these restrictions must be renormalized, to be defined on the cube I^k with sides of length 1. In symbols,

$$S(\Gamma) = \Sigma\Gamma_{\varepsilon_1,\ldots,\varepsilon_k},$$

the sum over all 2^k choices of $\varepsilon_i = 0$ or 1, and $\Gamma_{\varepsilon_1,\ldots,\varepsilon_k}$ is the k-cube defined by the formula

$$\Gamma_{\varepsilon_1,\ldots,\varepsilon_k}(t_1, \ldots, t_k) = \Gamma(\tfrac{1}{2}(t_1+\varepsilon_1), \ldots, \tfrac{1}{2}(t_k+\varepsilon_k)).$$

This is extended by linearity to give a homomorphism S: $C_kX \to C_kX$, i.e., by defining $S(\Sigma n_\alpha \Gamma_\alpha) = \Sigma n_\alpha S(\Gamma_\alpha)$. If the boundary of $S(\Gamma)$ is calculated, all the terms corresponding to inner faces cancel, and one gets the result of subdividing the boundary of Γ. In symbols,

$$\partial \circ S = S \circ \partial, \tag{23.4}$$

as homomorphisms from C_kX to $C_{k-1}X$.

Exercise 23.5. Verify this formula.

We can iterate this subdivision operation, defining for any k-chain c new k-chains $S(c)$, $S^2(c) = S(S(c))$, $S^3(c) = S(S(S(c)))$, and so on.

Lemma 23.6. *If X is a union of two open sets U and V, and c is a k-chain on X, then, for sufficiently large p, $S^p(c)$ can be written as a sum $c_1 + c_2$ where c_1 is a k-chain on U and c_2 is a k-chain on V.*

Proof. This is an immediate consequence of the Lebesgue lemma, since, if $\Gamma: I^k \to X$ is a k-cube, the image of each k-cube appearing in $S^p(\Gamma)$ is the image of a subcube of I^k with sides of length $1/2^p$. □

To use this construction and lemma, we want to know that, if z is a k-cycle, then $S(z)$ is a k-cycle defining the same homology class as z. For this we proceed as follows. Let α: $[0, 1] \to [0, 1]$ be defined by the formula

$$\alpha(t) = \begin{cases} 2t, & 0 \le t \le \tfrac{1}{2}, \\ 1, & \tfrac{1}{2} \le t \le 1. \end{cases}$$

If Γ is a k-cube in X, define a k-cube $A(\Gamma)$ by the formula

$$A(\Gamma)(t_1, \ldots, t_k) = \Gamma(\alpha(t_1), \ldots, \alpha(t_k)).$$

Note that the "first corner" $A(\Gamma)_{0,\ldots,0}$ of this k-cube is Γ, and all the other $A(\Gamma)_{\varepsilon_1,\ldots,\varepsilon_k}$ are degenerate. Extend this by linearity as usual to

a homomorphism A: $C_kX \to C_kX$. (If $k = 0$, define A to be the identity map.) By the observation just made, we have

$$S \circ A = I, \tag{23.7}$$

where I: $C_kX \to C_kX$ is the identity map. Now define, for a k-cube Γ, a $(k+1)$-cube $H(\Gamma)$ by the formula

$$H(\Gamma)(s, t_1, \ldots, t_k) = \Gamma((1-s)\alpha(t_1) + st_1, \ldots, (1-s)\alpha(t_k) + st_k),$$

and extend by linearity to a homomorphism H: $C_kX \to C_{k+1}X$. (If $k = 0$, set $H = 0$.) Note that $\partial_1^0(H(\Gamma)) = A(\Gamma)$ and $\partial_1^1(H(\Gamma)) = \Gamma$. From this one sees that

$$\partial \circ H + H \circ \partial = I - A, \tag{23.8}$$

as homomorphisms from C_kX to C_kX.

Exercise 23.9. Verify this formula.

For each $p \geq 1$, define a homomorphism R_p: $C_kX \to C_{k+1}X$ by the formula

$$R_p = S \circ H \circ (I + S + S^2 + \ldots + S^{p-1}).$$

Then we have, for all $p \geq 1$, the identity

$$\partial \circ R_p + R_p \circ \partial = S^p - I. \tag{23.10}$$

In fact, this is a formal calculation, following from (23.4), (23.7), and (23.8), as follows. When $p = 1$, $R_1 = S \circ H$, and

$$\begin{aligned} \partial \circ R_1 + R_1 \circ \partial &= \partial \circ S \circ H + S \circ H \circ \partial \\ &= S \circ \partial \circ H + S \circ H \circ \partial \\ &= S \circ (\partial \circ H + H \circ \partial) = S \circ (I - A) \\ &= S - S \circ A = S - I. \end{aligned}$$

For $p > 1$, $R_p = R_1 \circ S_p$, where $S_p = I + S + \ldots + S^{p-1}$. We use the case $p = 1$ in the form $\partial \circ R_1 = S - I - R_1 \circ \partial$, and we use the fact that ∂ commutes with S_p by (23.4), together with the identity $S^p - I = (S - I) \circ S_p$. Calculating, we have

$$\begin{aligned} \partial \circ R_p + R_p \circ \partial &= \partial \circ R_1 \circ S_p + R_p \circ \partial \\ &= (S - I - R_1 \circ \partial) \circ S_p + R_p \circ \partial \\ &= (S - I) \circ S_p + (-R_1 \circ \partial) \circ S_p + R_p \circ \partial \\ &= S^p - I - R_1 \circ S_p \circ \partial + R_p \circ \partial \\ &= S^p - I - R_p \circ \partial + R_p \circ \partial = S^p - I, \end{aligned}$$

as asserted.

Now suppose that $X = U \cup V$. The definition of the boundary homomorphism from $H_k(U \cup V)$ to $H_{k-1}(U \cap V)$ depends on the following lemma:

Lemma 23.11. *(a) Any homology class in H_kX can be represented by a cycle z on X of the form $z = c_1 + c_2$, where c_1 is a k-chain on U and c_2 is a k-chain on V. (b) The $(k-1)$-chain $\partial c_1 = -\partial c_2$ is a cycle on $U \cap V$, and its homology class in $H_{k-1}(U \cap V)$ is independent of choice of c_1 and c_2.*

Proof. For (a), take any cycle c that represents the homology class. By Lemma 23.6, for some $p \geq 1$, the chain S^pc can be written as the sum of a chain c_1 on U and a chain c_2 on V. By (23.10),

$$S^pc - c \;=\; \partial(R_p(c)) + R_p(\partial(c)) = \partial(R_p(c)),$$

from which it follows that $z = S^pc$ is a cycle representing the same homology class as c.

For (b), suppose $z' = c_1' + c_2'$ is another representative of the same form for the same homology class. There is a $(k+1)$-chain w on X with $\partial(w) = z' - z$. By Lemma 23.6, there is a $p \geq 1$ such that $S^p(w)$ can be written as a sum of a chain on U and a chain on V. Applying (23.10) to the chain $\partial(w)$, we have

$$\begin{aligned} z' - z \;=\; \partial(w) \;&=\; S^p(\partial(w)) - R_p(\partial(\partial(w))) - \partial(R_p(z' - z)) \\ &=\; \partial(S^p(w)) - \partial(R_p(z' - z)) \;=\; \partial(S^p(w) - R_p(z' - z)). \end{aligned}$$

From the formula for R_p it follows that R_p takes a chain on U to a chain on U and a chain on V to a chain on V. We know that $z' - z$ is a sum of a chain on U and a chain on V, and it follows that $R_p(z' - z)$ is also. It follows that there are $(k+1)$-chains y_1 and y_2 on U and V such that $S^p(w) - R_p(z' - z) = y_1 + y_2$, so

$$z' - z \;=\; \partial(y_1 + y_2).$$

This means that we have an equality of k-chains

$$c_1' - c_1 - \partial(y_1) \;=\; -(c_2' - c_2 - \partial(y_2)),$$

the left side of which is a chain on U and the right side is a chain on V. This chain, denoted x, is a chain on $U \cap V$, and it follows that

$$\partial(x) \;=\; \partial(c_1') - \partial(c_1),$$

so the cycles $\partial(c_1')$ and $\partial(c_1)$ differ by a boundary on $U \cap V$, as asserted. □

Define $\partial\colon H_kX \to H_{k-1}(U \cap V)$ by taking the homology class $[z]$ of

a cycle z of the form $c_1 + c_2$ with c_1 and c_2 chains on U and V, to the homology class $[\partial(c_1)]$ of the cycle $\partial(c_1) = -\partial(c_2)$ on $U \cap V$. It follows from Lemma 23.11 that this definition makes sense. The proof that it is a homomorphism, and that the resulting Mayer–Vietoris sequence is exact, is precisely the same as in Chapter 10, so will not be repeated; the general algebra for this will be described in §24a.

In fact, the above argument shows something more. Let X be any space, and $\mathcal{U} = \{U_\alpha : \alpha \in \mathcal{A}\}$ any open covering of X. Let $C_k(X)^{\mathcal{U}}$ be the subgroup of $C_k(X)$ consisting of linear combinations of nondegenerate k-cubes $\Gamma: I^k \to X$ such that the image of Γ is contained in one (or more) of the open sets U_α. These cubes are said to be *small* with respect to $\mathcal{U}$. The boundary operator maps $C_k(X)^{\mathcal{U}}$ to $C_{k-1}(X)^{\mathcal{U}}$, so we can define the corresponding homology groups $H_k(X)^{\mathcal{U}}$. There is a natural map from $H_k(X)^{\mathcal{U}}$ to $H_k(X)$. The following proposition says that we can always calculate homology by using chains that are small with respect to any convenient covering.

Proposition 23.12. *The natural map $H_k(X)^{\mathcal{U}} \to H_k(X)$ is an isomorphism.*

Exercise 23.13. Use (23.4), (23.7), (23.8), and (23.10) to prove this proposition.

This proposition also gives a more concise way to construct the Mayer–Vietoris exact sequence:

Exercise 23.14. For $\mathcal{U} = \{U, V\}$, construct a homomorphism from $H_k(X)^{\mathcal{U}}$ to $H_{k-1}(U \cap V)$, and show that there is a long exact sequence

$$\ldots \xrightarrow{\partial} H_k(U \cap V) \xrightarrow{-} H_kU \oplus H_kV \xrightarrow{+} H_k(X)^{\mathcal{U}} \xrightarrow{\partial} H_{k-1}(U \cap V) \to \ldots .$$

Combine with the proposition to get the full Mayer–Vietoris sequence.

The following is a useful general consequence of the Mayer–Vietoris sequence.

Exercise 23.15. Suppose a space X is a union of some open sets $U_1, \ldots, U_p$ such that all homology groups $H_k(Y)$ vanish for any intersection $Y = U_{i_1} \cap \ldots \cap U_{i_r}$ of these open sets and all $k > 0$. (a) Show that $H_k(X) = 0$ for $k \geq p$. (b) If, in addition, each intersection Y is connected, and $p \geq 2$, show that $H_{p-1}(X) = 0$. (c) Finally, if each intersection Y is connected and nonempty, show that $H_k(X) = 0$ for all $k > 0$.

23c. Spheres and Degree

We saw that S^n is simply connected if $n \geq 2$, so $H_1 S^n = 0$. With Mayer–Vietoris, one can calculate the homology groups of all spheres S^n. One can cover S^n by two open sets U and V each homeomorphic to open disks in $\mathbb{R}^n$, whose intersection is homeomorphic to $S^{n-1} \times I^{\circ}$ for an open interval I°. Mayer–Vietoris gives an isomorphism

$$\partial\colon H_k(S^n) \xrightarrow{\cong} H_{k-1}(S^{n-1} \times I^{\circ}) \;\cong\; H_{k-1}(S^{n-1})$$

for all $k > 1$. From this one sees that

$$H_k(S^n) \;\cong\; \begin{cases} \mathbb{Z} & \text{if } k = 0 \text{ or } n, \\ 0 & \text{otherwise.} \end{cases}$$

Exercise 23.16. Show that the complements of the south and north poles $(0, \ldots, 0, -1)$ and $(0, \ldots, 0, 1)$ satisfy the conditions for U and V, and use Mayer–Vietoris to complete the proof of this calculation.

Equipped with homology groups, the definition of the degree of a continuous map $f\colon S^n \to S^n$ is easy: Since $H_n S^n \cong \mathbb{Z}$, the induced map $f_*\colon H_n S^n \to H_n S^n$ is multiplication by an integer, and this integer is defined to be the degree of f, denoted $\deg(f)$. Equivalently, if $[z]$ is a generator for $H_n S^n$, $\deg(f)$ is the integer such that $f_*([z]) = \deg(f) \cdot [z]$. (Note that using the other generator $-[z]$ leads to the same degree.)

Exercise 23.17. (a) Show that homotopic maps from S^n to S^n have the same degree. (b) Show that if $f\colon S^n \to S^n$ extends to a continuous map from D^{n+1} to S^n, then $\deg(f) = 0$. (c) Show that if f and g are maps from S^n to S^n, then $\deg(g \circ f) = \deg(g) \cdot \deg(f)$. (d) Show that for any integer d and any $n \geq 1$ there are maps $f\colon S^n \to S^n$ of degree d.

In fact the converses of (a) and (b) of the preceding exercise are true, but more difficult.

Having the notion of degree, we can define the generalization of winding number: the *engulfing number* $W(f, P)$ of a continuous map $f\colon S^{n-1} \to \mathbb{R}^n \setminus \{P\}$ *around* P. This can be defined to be the degree of the map that follows f by projection onto a sphere around P, i.e., define $W(f, P)$ to be the degree of the map

$$S^{n-1} \to S^{n-1}, \qquad x \mapsto \frac{f(x) - P}{\|f(x) - P\|}.$$

Problem 23.18. (a) Show that, as a function of P, this number is constant on connected components of $\mathbb{R}^n \setminus f(S^{n-1})$. (b) State and prove an n-dimensional analogue of the dog-on-a-leash theorem. (c) Show that for $f\colon S^n \to \mathbb{R}^n \setminus \{0\}$ differentiable, this definition agrees with that in Problem 22.23.

Similarly, one can define the *local degree* of a continuous mapping $f\colon U \to V$ between open sets in $\mathbb{R}^n$ at a point P in U, provided there is a neighborhood U_P of P such that no other point of U_P has the same image as P. This is the degree of the mapping

$$S^{n-1} \to S^{n-1}, \qquad x \mapsto \frac{f(P+rx)-f(P)}{\|f(P+rx)-f(P)\|},$$

for any positive r so that U_P contains the ball of radius r around P. This can be used for example to define the notion of a homeomorphism from $\mathbb{R}^n$ to $\mathbb{R}^n$ (or S^n to S^n, or X to X for any oriented manifold) being *orientation preserving* or *orientation reversing*, according as the local degree at any point is $+1$ or -1.

Exercise 23.19. (a) Show that this local degree is a continuous function of the point, so the notion of orientation preserving or reversing is well defined. (b) Show that a homeomorphism of S^n is orientation preserving or reversing according as its degree is $+1$ or -1. Show that a homeomorphism of $\mathbb{R}^n$ extends to a homeomorphism of $S^n = \mathbb{R}^n \cup \{\infty\}$, whose degree therefore determines whether the original map is orientation preserving. (c) For a diffeomorphism, show that the local degree is given by the sign of the determinant of the Jacobian.

With the concept of degree, the following assertions of Borsuk and Brouwer are proved just as before, and the proofs are left as exercises.

Theorem 23.20. (1) *There is no retraction from D^{n+1} onto S^n.*

(2) *Any continuous mapping from a closed disk D^n to itself must have a fixed point.*

Exercise 23.21. Generalize the results of Exercises 4.9–4.17.

Problem 23.22. Show that the degree of the antipodal map from S^n to S^n is 1 if n is odd and -1 if n is even. In particular, the antipodal map is not homotopic to the identity if n is even.

Problem 23.23. (a) Show that no even-dimensional sphere can have a nowhere vanishing vector field. (b) Construct on every odd-dimensional sphere a nowhere vanishing vector field.

To extend the results about antipodal mappings, we need the following generalization of Borsuk's Lemma 4.20. As there, we denote by $P^* = -P$ the antipode of a point P in S^n.

Theorem 23.24. *Let* $f\colon S^n \to S^n$ *be a continuous map.*

(a) If $f(P^*) = f(P)^*$ *for all* P *in* S^n*, then the degree of* f *is odd.*
(b) If $f(P^*) = f(P)$ *for all* P *in* S^n*, then the degree of* f *is even.*

The proof of this requires some new ideas, and is postponed to Appendix E. Assuming this theorem, we can draw the expected consequences:

Corollary 23.25. *(a) If* $m < n$*, there is no continuous mapping* $f\colon S^n \to S^m$ *such that* $f(P^*) = f(P)^*$ *for all* P *in* S^n.

(b) Any continuous mapping $f\colon S^n \to \mathbb{R}^n$ *must map some pair of antipodal points to the same point.*

(c) An open set in $\mathbb{R}^n$ *cannot be homeomorphic to an open set in* $\mathbb{R}^m$ *if* $n \neq m$.

(d) It is impossible to cover S^n *with* $n+1$ *closed sets, none of which contains a pair of antipodal points.*

Exercise 23.26. Prove this corollary.

Problem 23.27. Let $f\colon S^n \to S^n$ be continuous. (a) If $f(P^*) \neq f(P)$ for all P, show that $\deg(f)$ is odd. (b) If $f(P^*) \neq f(P)^*$ for all P, show that $\deg(f)$ is even. (c) Show that the only nontrivial group that can act freely on an even-dimensional sphere is the group with two elements.

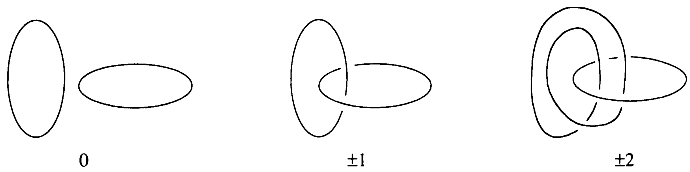
Exercise 23.28. Show that if $n > 1$ any continuous mapping from S^n to $\mathbb{R}P^n$ must map some pair of antipodal points to the same point.

Exercise 23.29. Prove that if $n+1$ bounded measurable objects are given in $\mathbb{R}^n$, then there is a hyperplane that cuts each in half.

Exercise 23.30. (a) State and prove n-dimensional analogues of Exercises 4.24–4.31 and 4.34–4.39. (b) Define the index of a vector field on an open set in $\mathbb{R}^n$ at an isolated singular point, and state and prove the n-dimensional analogues of Proposition 7.5 and its corollaries.

Exercise 23.31. Use Mayer–Vietoris to compute the homology groups of a torus $S^1 \times S^1$, or of any product $S^m \times S^n$.

Homology can be used to define a notion of degree in many other contexts. Here is an important illustration. A *link* in $\mathbb{R}^3$ is a disjoint union of knots. Equivalence is defined just as for knots. A link can be nontrivial even if all the knots occurring in it are trivial. There is a *linking number* that measures how many times two knots intertwine with each other.



Suppose K and L are disjoint knots. Define a mapping

$$F\colon K \times L \to S^2, \qquad x \times y \mapsto \frac{x-y}{\|x-y\|},$$

which assigns to the pair (x, y) the direction from x to y. This mapping F determines a homomorphism $F_*\colon H_2(K \times L) \to H_2(S^2)$. Both of these homology groups are isomorphic with $\mathbb{Z}$. Choosing orientations of each identifies them with $\mathbb{Z}$, and then F_* is multiplication by some integer, which is defined to be the linking number $l(K, L)$. If a standard orientation is fixed for S^2, the sign of $l(K, L)$ depends on orientations chosen for K and L; it changes sign if either of these orientations are changed. The linking number also changes sign if the roles of K and L are reversed. Note that if K and L are far apart, then F will not be surjective, so this linking number is zero.

Exercise 23.32. (a) Show that the linking number of the two circles in Exercise 22.7 is ± 1, and therefore the linking number of two fibers of the Hopf mapping of Exercise 22.17 is ± 1. (b) Show that the intersection of a small sphere S^3 around a singularity of a plane curve that is a node (see §20c) is two circles whose linking number is ± 1.

It is a fact (see §24c) that the top homology group H_nX of an oriented n-manifold X is $\mathbb{Z}$ (the orientation determining a choice of generator). It follows that any continuous map $f\colon X \to Y$ between oriented n-manifolds has a degree.

Problem 23.33. (a) Show that if X is a compact oriented surface, then $H_2X \cong \mathbb{Z}$. (b) If X is a compact nonorientable surface, show that $H_2X = 0$. (c) If $f\colon X \to Y$ is a nonconstant analytic map between compact Riemann surfaces, show that the degree defined by homology is the same as the number of sheets of the corresponding branched covering.

23d. Generalized Jordan Curve Theorem

There is a vast generalization of the Jordan curve theorem to higher dimensions. This can be stated as follows:

Theorem 23.34. *If $X \subset S^n$ is homeomorphic to a sphere S^m, then $m \le n$, and $m < n$ unless $X = S^n$. If $m < n$, the homology groups of the complement are*

$$H_k(S^n \setminus X) \cong \begin{cases} \mathbb{Z} \oplus \mathbb{Z} & \text{if } m = n-1 \text{ and } k = 0, \\ \mathbb{Z} & \text{if } m < n-1 \text{ and } k = 0 \text{ or } k = n-1-m, \\ 0 & \text{otherwise.} \end{cases}$$

In particular, the complement has two components if $m = n - 1$, and one if $m < n - 1$.

The essential point of this theorem is the assertion that the homology of the complement is the same for all embeddings of S^m in S^n. The proof follows the pattern for the Jordan Curve Theorem in the plane very closely, using the full Mayer–Vietoris theorem. We discuss only the new features, leaving details to the reader. First we have the analogous result for embeddings of cubes in S^n.

Proposition 23.35. *If $X \subset S^n$ is homeomorphic to I^m, then $S^n \setminus X$ is connected and $H_k(S^n \setminus X) = 0$ for all $k > 0$.*

Proof. This is by induction on m, the case $m = 0$ being clear since the complement of a point is homeomorphic to $\mathbb{R}^n$, so contractible. For $m > 0$ write I^m as the union of two halves whose intersection is homeomorphic to I^{m-1}; this makes X a union of two subspaces A and B. Applying Mayer–Vietoris to $U = S^n \setminus A$ and $V = S^n \setminus B$, and knowing about $U \cup V$ by induction, we have

$$0 = H_{k+1}(U \cup V) \to H_k(U \cap V) \to H_kU \oplus H_kV.$$

From this it follows that if z is a k-cycle on $S^n \setminus X$ that is not a bound-

ary, then it is not a boundary on $S^n \setminus A$ or $S^n \setminus B$. Continuing to cut the cubes in half, passing to the limit as in Chapter 5, we find that z is not a boundary on $S^n \setminus \{x\}$ for x a point, from which the conclusion follows easily. □

To prove the theorem, also by induction on m, write X as the union of two closed sets A and B homeomorphic to the upper and lower hemispheres of the sphere S^m. Each of A and B is homeomorphic to I^m, and $A \cap B$ is homeomorphic to S^{m-1}. Applying Mayer–Vietoris and the proposition to $U = S^n \setminus A$ and $V = S^n \setminus B$, we get

$$0 \to H_{k+1}(S^n \setminus A \cap B) \to H_k(S^n \setminus X) \to 0$$

if $k > 0$, and

$$0 \to H_1(S^n \setminus A \cap B) \to H_0(S^n \setminus X) \to H_0 U \oplus H_0 V \to H_0(S^n \setminus A \cap B) \to 0.$$

We know about $S^n \setminus A \cap B$ by induction, so the first display computes $H_k(S^n \setminus X)$ for $k > 0$. If $m < n - 1$, $H_1(S^n \setminus A \cap B) = 0$, and the second gives

$$0 \to H_0(S^n \setminus X) \to \mathbb{Z} \oplus \mathbb{Z} \to \mathbb{Z} \to 0,$$

from which it follows easily that $H_0(S^n \setminus X) \cong \mathbb{Z}$. If $m = n - 1$, then $H_1(S^n \setminus A \cap B) \cong \mathbb{Z}$, and from

$$0 \to \mathbb{Z} \to H_0(S^n \setminus X) \to \mathbb{Z} \oplus \mathbb{Z} \to \mathbb{Z} \to 0$$

we find similarly that $H_0(S^n \setminus X) \cong \mathbb{Z} \oplus \mathbb{Z}$. If $m = n$, from

$$0 \to H_0(S^n \setminus X) \to \mathbb{Z} \oplus \mathbb{Z} \to \mathbb{Z} \oplus \mathbb{Z} \to 0$$

we see that $H_0(S^n \setminus X) = 0$, so $X = S^n$. From this it follows that no larger sphere can be embedded in S^n, since an n-dimensional subsphere would already map to the whole S^n. □

Exercise 23.36. State and prove analogous results for $X \subset \mathbb{R}^n$ homeomorphic to a sphere.

Exercise 23.37. If $F\colon D^n \to \mathbb{R}^n$ is continuous and one-to-one, show that $\mathbb{R}^n \setminus F(S^{n-1})$ has two connected components, one the image of the interior of D^n, the other the complement of $F(D^n)$. Deduce the *invariance of domain*: if U is open in $\mathbb{R}^n$, and $F\colon U \to \mathbb{R}^n$ is a one-to-one continuous mapping, then $F(U)$ is open. Prove the *invariance of dimension*: open sets in $\mathbb{R}^n$ and $\mathbb{R}^m$ cannot be homeomorphic if $n \neq m$.

It is also true that if $F\colon S^n \to \mathbb{R}^{n+1}$ is an embedding, then its en-

gulfing number around points in the two components of the complements is ± 1 for the bounded component and 0 for the unbounded component (see Proposition 5.20), but this requires more machinery.

Exercise 23.38. Suppose A and B are disjoint closed subsets of S^n, $n > 1$, and two points are given in the complement of $A \cup B$. Show that if A and B do not separate these points, neither does $A \cup B$.

The Mayer–Vietoris sequence can also be used to show that a compact nonorientable surface cannot be topologically embedded in $\mathbb{R}^3$. We sketch a proof in the following problems:

Problem 23.39. (a) Show that if $X \subset \mathbb{R}^3$ is homeomorphic to a Moebius band, then $H_k(\mathbb{R}^3 \setminus X) \cong \mathbb{Z}$ if $k = 0$ or 1, and $H_k(\mathbb{R}^3 \setminus X) = 0$ otherwise. (b) With X as in (a), if $Y \subset X$ corresponds to the boundary circle, show that the map

$$H_1(\mathbb{R}^3 \setminus X) \to H_1(\mathbb{R}^3 \setminus Y) \cong \mathbb{Z}$$

determined by the inclusion of $\mathbb{R}^3 \setminus X$ in $\mathbb{R}^3 \setminus Y$ takes a generator of the first group to twice a generator of the second.

Suppose a subspace X of $\mathbb{R}^3$ were homeomorphic to the projective plane. Write X as a union of a space A homeomorphic to a Moebius band and B homeomorphic to a disk, with $A \cap B$ homeomorphic to the boundary circle of each. Then with $U = \mathbb{R}^3 \setminus A$, $V = \mathbb{R}^3 \setminus B$, Mayer–Vietoris and Exercise 23.36 give an exact sequence

$$H_1(\mathbb{R}^3 \setminus A) \oplus 0 \to H_1(\mathbb{R}^3 \setminus A \cap B) \to H_0(\mathbb{R}^3 \setminus X).$$

By the preceding problem, the image of a generator of $H_1(\mathbb{R}^3 \setminus A \cap B)$ maps to an element α in $H_0(\mathbb{R}^3 \setminus X)$ that is nonzero, but $2 \cdot \alpha = 0$. However, we know that the 0th homology group of any space is a free abelian group, which has no such element.

Problem 23.40. Show similarly that none of the nonorientable compact surfaces can be embedded in $\mathbb{R}^3$.

The following problem generalizes two of the main results of Chapters 6 and 9:

Problem 23.41. Let U be an open subset of $\mathbb{R}^n$. (a) Show that two classes in $H_{n-1}U$ are equal if and only if they map to equal classes in $H_{n-1}(\mathbb{R}^n \setminus \{P\})$ for all P not in U. (b) Show that if the complement of U in $S^n = \mathbb{R}^n \cup \{\infty\}$ is a disjoint union of $m+1$ compact connected sets, then $H_{n-1}U$ is a free abelian group on m generators.

CHAPTER 24

Duality

24a. Two Lemmas from Homological Algebra

We frequently want to compare different homology and cohomology groups, when we have exact Mayer–Vietoris sequences for each, and maps between them. Assuming that most of the maps are isomorphisms, we want to deduce that the others are as well. There is a general algebraic fact that can be used for this:

Lemma 24.1 (Five-Lemma). *Given a commutative diagram*

$$\begin{array}{ccccccccc} A & \longrightarrow & B & \longrightarrow & C & \longrightarrow & D & \longrightarrow & E \\ \downarrow & & \downarrow & & \downarrow & & \downarrow & & \downarrow \\ A' & \longrightarrow & B' & \longrightarrow & C' & \longrightarrow & D' & \longrightarrow & E' \end{array}$$

of abelian groups, such that the rows are exact sequences, and all the vertical maps but the middle one are isomorphisms, then the middle map from C to C' must also be an isomorphism.

The proof is by a "diagram chase," which is much easier and enjoyable to do for oneself than to follow when someone else does it. Here is how to show that the map is one-to-one. If c in C maps to 0 in C', then its image in D maps to 0 in D' (by commutativity of the diagram), so c maps to 0 in D (since $D \rightarrow D'$ is injective), so c comes from some element b in B (by exactness of the top row). The element b maps to an element b' in B' that maps to 0 in C' (why?), that

therefore comes from an element a' in A'. This element a' comes from some element a in A, and this element a must map to b since they have the same images in B'. Since a maps to 0 in C, and b maps to c, c must be 0.

Exercise 24.2. (a) Prove similarly that $C \to C'$ is surjective. (b) Show that the five-lemma is also valid under the following weaker assumptions, still assuming the rows are exact: (i) each square either commutes or commutes up to sign, i.e., the composite going around one way is plus or minus the composite going around the other way; and (ii) the maps $B \to B'$ and $D \to D'$ are isomorphisms, $A \to A'$ is surjective, and $E \to E'$ is injective.

If you like this diagram chasing, there is a general process that constructs long exact sequences, which can be used to construct all the Mayer–Vietoris sequences we have seen. For this, one has a commutative diagram of abelian groups

$$
\begin{array}{ccccccccc}
\downarrow & & \downarrow & & \downarrow & & \downarrow & & \downarrow \\
0 & \longrightarrow & C_{k+1}' & \longrightarrow & C_{k+1} & \longrightarrow & C_{k+1}'' & \longrightarrow & 0 \\
\downarrow & & \downarrow & & \downarrow & & \downarrow & & \downarrow \\
0 & \longrightarrow & C_{k}' & \longrightarrow & C_{k} & \longrightarrow & C_{k}'' & \longrightarrow & 0 \\
\downarrow & & \downarrow & & \downarrow & & \downarrow & & \downarrow \\
0 & \longrightarrow & C_{k-1}' & \longrightarrow & C_{k-1} & \longrightarrow & C_{k-1}'' & \longrightarrow & 0 \\
\downarrow & & \downarrow & & \downarrow & & \downarrow & & \downarrow \\
0 & \longrightarrow & C_{k-2}' & \longrightarrow & C_{k-2} & \longrightarrow & C_{k-2}'' & \longrightarrow & 0, \\
\downarrow & & \downarrow & & \downarrow & & \downarrow & & \downarrow
\end{array}
$$

where the rows are exact, and the composite of any two successive maps in the columns is zero. One says that the columns are *chain complexes*. The diagram is abridged to saying one has a *short exact sequence of chain complexes*

$$0 \to C_*' \to C_* \to C_*'' \to 0.$$

For each chain complex (column) one can form *homology groups*. For the center column,

$$H_k(C_*) = Z_k(C_*)/B_k(C_*),$$

where $Z_k(C_*) = \text{Kernel}(C_k \to C_{k-1})$ are the *k-cycles*, and $B_k(C_*) = \text{Image}(C_{k+1} \to C_k)$ are the *k-boundaries*. Similarly for the other two columns. There are maps from $H_k(C_*')$ to $H_k(C_*)$ and from $H_k(C_*)$ to $H_k(C_*'')$, determined by the horizontal maps in the diagram. For example, the map from C_k' maps $Z_k(C_*')$ to $Z_k(C_*)$ and $B_k(C_*')$ to $B_k(C_*)$, so it determines a homomorphism on the quotient group. The interesting maps are *boundary homomorphisms*

$$\partial\colon H_k(C_*'') \to H_{k-1}(C_*').$$

To define these, take a representative z'' in $Z_k(C_*'')$ of a class in $H_k(C_*'')$. Choose an element c in C_k that maps onto z''. Let $\bar{c}$ be the image of c in C_{k-1}. Then $\bar{c}$ maps to 0 in C_{k-1}'', since it has the same image there as z'' does, and z'' is a cycle. So $\bar{c}$ comes from an element c' in C_{k-1}'.

Exercise 24.3. Show that this element c' is a $(k-1)$-cycle, and its homology class in $H_{k-1}(C_*')$ is independent of choices of the representative z'' and the element c that maps onto z''.

The homology class of c' is defined to be the boundary of the homology class of z'': $\partial([z'']) = [c']$. One checks easily that ∂ is a homomorphism of abelian groups.

Proposition 24.4. *The resulting sequence*

$$\ldots \to H_{k+1}(C_*'') \to H_k(C_*') \to H_k(C_*) \to H_k(C_*'') \to H_{k-1}(C_*'') \to \ldots$$

is exact.

The proof is some more diagram chasing, which again we leave as an exercise. There is a similar result when the vertical maps in the diagram go up rather than down. Usually then the indexing is by upper indices, so we have "cochain complexes"

$$C^*\colon \qquad \ldots \to C^{k-1} \to C^k \to C^{k+1} \to \ldots .$$

A short exact sequence $0 \to C^{*\prime} \to C^* \to C^{*\prime\prime} \to 0$ determines a long exact sequence of their *cohomology groups*

$$\ldots \to H^{k-1}(C^{*\prime\prime}) \to H^k(C^{*\prime}) \to H^k(C^*) \\ \to H^k(C^{*\prime\prime}) \to H^{k+1}(C^{*\prime}) \to \ldots .$$

Let us see how some of the Mayer–Vietoris sequences we have seen earlier fall out of this formalism. For example, if X is a $\mathscr{C}^\infty$ manifold, and C^kX denotes the vector space of $\mathscr{C}^\infty$ k-forms on X, and

U and V are open sets in X, there is an exact sequence

$$0 \to C^*(U \cup V) \to C^*(U) \oplus C^*(V) \to C^*(U \cap V) \to 0$$

of cochain complexes. The first map takes a form ω on $U \cup V$ to the pair $(\omega|_U, \omega|_V)$, and the second takes a pair (ω_1, ω_2) to the difference $\omega_1|_{U\cap V} - \omega_2|_{U\cap V}$. The exactness of this sequence is clear from the definitions except for the surjectivity of the second map, and that was proved using a partition of unity in Chapter 10. The Mayer–Vietoris sequence then results immediately from Proposition 24.4.

Similarly, for cohomology with compact support, one has a short exact sequence of cochain complexes

$$0 \to C_c^*(U \cap V) \to C_c^*(U) \oplus C_c^*(V) \to C_c^*(U \cup V) \to 0,$$

where the first map takes a form ω on $U \cap V$ to the pair $(\omega^U, -\omega^V)$, where ω^U denotes the extension by 0 from $U \cap V$ to U, and similarly for V; the second map takes (ω_1, ω_2) to $\omega_1{}^{U\cup V} + \omega_2{}^{U\cup V}$. Again, exactness follows from a partition of unity argument, and the Mayer–Vietoris exact sequence results.

For homology, if a space X is a union of open sets U and V, let $C_k(X)^{\mathcal{U}}$ denote the k-chains that are small with respect to the covering $\mathcal{U} = \{U, V\}$ of X. Then there is an exact sequence

$$0 \to C_*(U \cap V) \to C_*(U) \oplus C_*(V) \to C_*(X)^{\mathcal{U}} \to 0$$

of chain complexes, the first taking a chain on $U \cap V$ to the pair consisting of its images on U and on V, and the second taking a pair to the difference of their images on X. The exactness is immediate, giving an exact sequence

$$\ldots \to H_{k+1}(C_*(X)^{\mathcal{U}}) \to H_k(U \cap V) \to H_k(U) \oplus H_k(V)$$
$$\to H_k(C_*(X)^{\mathcal{U}}) \to H_{k-1}(U \cap V) \to \ldots \ .$$

To complete the proof, one appeals to Proposition 23.12, which says that $H_k(C_*(X)^{\mathcal{U}}) \cong H_k(C_*(X)) = H_k(X)$.

Exercise 24.5. If $0 \to C' \to C \to C'' \to 0$ is an exact sequence of free abelian groups, and G is any abelian group, show that

$$0 \to \operatorname{Hom}(C'', G) \to \operatorname{Hom}(C, G) \to \operatorname{Hom}(C', G) \to 0$$

is also an exact sequence. If $0 \to C_*' \to C_* \to C_*'' \to 0$ is an exact sequence of complexes of free abelian groups, this gives an exact sequence $0 \to \operatorname{Hom}(C_*'', G) \to \operatorname{Hom}(C_*, G) \to \operatorname{Hom}(C_*', G) \to 0$ of cochain complexes, and hence an exact sequence of cohomology groups.

Exercise 24.6. Let C_* and C_*' be chain complexes with boundary maps denoted $\partial_k: C_k \to C_{k-1}$ and $\partial_k': C_k' \to C_{k-1}'$, respectively. Define a *map of chain complexes* $f_*: C_* \to C_*'$ to be a collection of homomorphisms $f_k: C_k \to C_k'$ that commute with the boundary maps. Show that such maps f_* determines homomorphisms from $H_k(C_*)$ to $H_k(C_*')$. Call two maps f_* and g_* *chain homotopic* if there is a collection of maps, $H_k: C_k \to C_{k+1}'$ such that

$$g_k - f_k = \partial_{k+1} \circ H_k + H_{k-1} \circ \partial_k$$

for all k. Show that f_* and g_* then determine the same maps from $H_k(C_*)$ to $H_k(C_*')$ for all k.

If $C_*' \to C_*$ is a map of chain complexes such that each $C_k' \to C_k$ is one-to-one, one can define C_k'' to be C_k/C_k', getting an exact sequence $0 \to C_*' \to C_* \to C_*/C_*' \to 0$, so a long exact homology sequence. For example, if Y is a subspace of a topological space X, then $C_*Y \to C_*X$ is one-to-one, so one can define a quotient chain complex C_*X/C_*Y. The homology groups of this complex are denoted $H_k(X, Y)$, and are called the *relative* homology groups. They fit in a long exact sequence

$$\ldots \to H_{k+1}(X, Y) \to H_k(Y) \to H_k(X) \to H_k(X, Y) \to H_{k-1}(Y) \to \ldots \quad .$$

For suitably nice spaces, these relative groups are isomorphic to the homology groups of the space obtained by collapsing (identifying) Y to a point. In many treatments of algebraic topology, these relative groups, and the above sequence, are used for calculation in most situations where we have used the Mayer–Vietoris sequence.

24b. Homology and De Rham Cohomology

In this section we want to prove that the De Rham cohomology groups H^kX of a manifold are dual to the homology groups H_kX, i.e., we want to construct an isomorphism

$$H^kX \xrightarrow{\cong} \operatorname{Hom}(H_kX, \mathbb{R}),$$

generalizing what we did for surfaces for $k = 1$. The idea is similar: one wants to integrate k-forms over k-cubes. This makes sense for differentiable k-cubes, but there is a problem of how to define this for continuous k-cubes that are not differentiable—a problem that we avoided for $k = 1$ by cutting up the path, locally writing the k-form

as the differential of a $(k-1)$-form μ, and evaluating μ over the endpoints of the subdivided path. For $k>1$, unless the restriction of the k-cube to its boundary is differentiable, this will not work. A more systematic procedure, that does work, is to show that the homology H_kX can be computed by using only differentiable cubes.

A cube $\Gamma\colon I^k \to X$ is a $\mathscr{C}^\infty$ *cube* if it extends to a $\mathscr{C}^\infty$ mapping on some neighborhood of the cube $I^k \subset \mathbb{R}^k$. Define $C_k^\infty X$ to be the free abelian group on the nondegenerate $\mathscr{C}^\infty$ k-cubes. The boundary ∂ from the k-chains to the $(k-1)$-chains takes $\mathscr{C}^\infty$ cubes to $\mathscr{C}^\infty$ cubes, so we can define the $\mathscr{C}^\infty$ homology groups $H_k^\infty X$ to be the quotient of the closed $\mathscr{C}^\infty$ k-chains modulo the subgroup consisting of boundaries of $\mathscr{C}^\infty$ $(k+1)$-chains. There is an obvious map

$$H_k^\infty X \to H_kX,$$

which, in the language of the preceding section, is given by the map of chain complexes $C_*^\infty X \to C_*X$. We will show that this is an isomorphism.

If ω is a k-form on X, and $\Gamma\colon I^k \to X$ is a $\mathscr{C}^\infty$ cube, we can define the integral of ω over Γ by

$$\int_\Gamma \omega = \int_{I^k} \Gamma^*(\omega),$$

where $\Gamma^*(\omega)$ is the pull-back form; a form on the cube can be written

$$f(x_1, \ldots, x_k)\, dx_1 \wedge dx_2 \wedge \ldots \wedge dx_k,$$

and the integral of such a form is the usual Riemann integral of the continuous function f on the cube.

Exercise 24.7. Prove "Stokes' theorem" in this context: if ω is a $(k-1)$-form, then

$$\int_\Gamma d\omega = \int_{\partial\Gamma} \omega.$$

From Stokes' theorem, it follows as in the case $k=1$ that there is a map

$$H^kX \to \operatorname{Hom}(H_k^\infty X, \mathbb{R}), \qquad \omega \mapsto \left[\Gamma \mapsto \int_\Gamma \omega\right].$$

We will see that this is also an isomorphism. Combining these two isomorphisms will give the duality we were after.

To prove these isomorphisms, we need to know that the groups $H_k^\infty X$ have many of the same properties as the purely topological groups $H_k X$. For example,

Exercise 24.8. (a) Show that a $\mathscr{C}^\infty$ mapping $f: X \to Y$ of manifolds determines functorial homomorphisms $f_*: H_k^\infty X \to H_k^\infty Y$. (b) If two maps from X to Y are homotopic by a $\mathscr{C}^\infty$ mapping $F: X \times I' \to Y$ (where I' is an open interval containing $[0, 1]$), then they determine the same homomorphisms. (c) Deduce that $H_k^\infty U = 0$ for $k > 0$, and $H_0^\infty U = \mathbb{Z}$ if U is a starshaped open set in $\mathbb{R}^n$.

Similarly, one has Mayer–Vietoris exact sequences just as for the groups $H_k X$, and compatible with the maps from groups H_k^∞ to the H_k. In fact, the same construction works in the $\mathscr{C}^\infty$ case, noting that the subdivision operators used to cut cubes into small pieces preserve $\mathscr{C}^\infty$ chains. One modification needs to be made in our proof, however, since the operator A we used in Chapter 23 used a function that is only piecewise differentiable.

Exercise 24.9. Change the function α used in §23b to a $\mathscr{C}^\infty$ function from $[0, 1]$ to $[0, 1]$ such that $\alpha(0) = 0$ and $\alpha(t) = 1$ if $t \geq 1/2$. With any such α, show that, for any chain Γ, $S \circ A(\Gamma) - \Gamma$ is a boundary. Use this to complete the proof of Mayer–Vietoris for these groups.

To prove these isomorphisms, we need a way to build up arbitrary manifolds out of simple pieces. The following general lemma will suffice for our purposes. Let us call an *open rectangle* in $\mathbb{R}^n$ an open rectangular solid with sides parallel to the axes, i.e., an open set of the form $(a_1, b_1) \times \ldots \times (a_n, b_n)$.

Lemma 24.10. *If X is an open set in $\mathbb{R}^n$, then X can be written as the union of two open sets U and V such that each of U and V and $U \cap V$ is a disjoint union of open sets, each of which is a finite union of open rectangles.*

Proof. Take compact sets $K_1 \subset K_2 \subset \ldots$ as in the Lemma A.20. Construct a sequence of open sets U_p as follows. Let U_1 be a finite union of rectangles covering K_1, with the closure of each contained in the interior of K_2. Let U_2 be a finite union of rectangles covering $K_2 \setminus \mathrm{Int}(K_1)$, with the closure of each contained in the interior of K_3. Inductively, let U_p be a finite union of rectangles that covers the compact set $K_p \setminus \mathrm{Int}(K_{p-1})$, the closure of each contained in the interior of

K_{p+1} and in the complement of K_{p-2}, and not meeting U_{p-2}. Now let U be the union of the union of all U_p with p even, and let V be the union of the union of all U_p with p odd. □

Lemma 24.11. *If X is a differentiable n-manifold, then X can be written as the union of two open sets U and V such that each of U and V and $U \cap V$ is a disjoint union of open sets, each of which is a finite union of open sets diffeomorphic to open sets in $\mathbb{R}^n$.*

Proof. The argument is the same. Remark A.21 shows that X is a union of compact sets $K_1 \subset K_2 \subset \ldots$ with the same properties. Then the preceding proof, with "rectangle" replaced by "open set diffeomorphic to an open set in $\mathbb{R}^n$" goes over without change. □

Theorem 24.12. *For any manifold X the natural maps $H_k^\infty X \to H_k X$ are isomorphisms.*

Proof. Let us write "$T(X)$" for the statement that the maps from $H_k^\infty X$ to $H_k X$ are isomorphisms for all k. There are three tools:

(1) $T(U)$ is true when U is an open rectangle in $\mathbb{R}^n$.
(2) If U and V are open in a manifold, and if $T(U)$, $T(V)$, and $T(U \cap V)$ are true, then $T(U \cup V)$ is true.
(3) If X is a disjoint union of open manifolds X_α, and each $T(X_\alpha)$ is true, then $T(X)$ is true.

With what we have seen, each of these is easy to prove. (1) follows from the fact that $H_k^\infty U$ and $H_k U$ vanish for $k > 0$, and both are naturally isomorphic to $\mathbb{Z}$ when $k = 0$, cf. Exercise 24.8. (2) follows from the fact that we have Mayer–Vietoris exact sequences for each, with compatible maps between them:

$$\begin{array}{ccccccccc}
H_k^\infty U\cap V & \longrightarrow & H_k^\infty U \oplus H_k^\infty V & \longrightarrow & H_k^\infty U \cup V & \longrightarrow & H_{k-1}^\infty U \cap V & \longrightarrow & H_{k-1}^\infty U \oplus H_{k-1}^\infty V \\
\downarrow & & \downarrow & & \downarrow & & \downarrow & & \downarrow \\
H_k U\cap V & \longrightarrow & H_k U \oplus H_k V & \longrightarrow & H_k U \cup V & \longrightarrow & H_{k-1} U \cap V & \longrightarrow & H_{k-1} U \oplus H_{k-1} V
\end{array}$$

so the five-lemma shows that the middle map is an isomorphism if the others are. (3) follows from the fact that to specify a class of either kind on X is equivalent to specifying a class on each X_α, with all but a finite number of these classes being zero (i.e., $H_k X$ is the direct sum of the groups $H_k X_\alpha$, and similarly for H_k^∞).

We can now use these tools to prove the theorem. We first show that $T(X)$ is true whenever $X \subset \mathbb{R}^n$ is a finite union of open rectangles. This is by induction on the number of rectangles. (1) takes care of

one rectangle, and if X is a union of p rectangles, let U be the union of $p - 1$ of them and let V be the other. Then $T(U)$ and $T(V)$ are true by induction, and $T(U \cap V)$ is true since $U \cap V$ is also a union of at most $p - 1$ rectangles, since the intersection of two rectangles is either empty or a rectangle. Then $T(X)$ is true by (2).

Next we show that $T(X)$ is true whenever X is an open set in $\mathbb{R}^n$. By Lemma 24.10 one can write X as a union of two open sets U and V, such that each of U and V and $U \cap V$ is a disjoint union of open sets, each of which is a finite union of open rectangles. Applying the preceding step and (3) we know that $T(U)$ and $T(V)$ and $T(U \cap V)$ are true, and by (2) again we know that $T(X)$ is true.

Note that since a diffeomorphism between manifolds determines an isomorphism between the corresponding groups, it follows that $T(X)$ is true for any set diffeomorphic to an open set in $\mathbb{R}^n$. The same inductive argument as for rectangles shows that $T(X)$ is true when X is a finite union of open sets, each diffeomorphic to an open set in $\mathbb{R}^n$. For the general case, Lemma 24.11 shows that any manifold X is a union of two open sets U and V such that each of U and V and $U \cap V$ is a disjoint union of open sets, each of which is diffeomorphic to a finite union of open sets in $\mathbb{R}^n$. By the last step and (3) again, $T(U)$ and $T(V)$ and $T(U \cap V)$ are true, and a final application of (2) shows that $T(X)$ is true. □

Theorem 24.13. *For any manifold X the natural maps from H^kX to $Hom(H_k^\infty X, \mathbb{R})$ are isomorphisms.*

Proof. The proof follows exactly the same format, with $T(X)$ being the statement that these maps are isomorphisms for all k. Once (1)–(3) are proved, in fact, the proof is identical. The proof of (1) is the same, and (3) follows from the fact that to specify a class of either kind is equivalent to specifying a class on each X_α (i.e., H^kX is the direct product of the groups H^kX_α, and similarly for $\mathrm{Hom}(H_k^\infty X, \mathbb{R})$). To prove (2), we need to compare the cohomology Mayer–Vietoris sequence with the dual of the Mayer–Vietoris sequence in homology. For brevity write $H_k^\infty X^*$ in place of $\mathrm{Hom}(H_k^\infty X, \mathbb{R})$. We have a diagram

$$\begin{array}{ccccccccc}
H^{k-1}U \oplus H^{k-1}V & \longrightarrow & H^{k-1}U \cap V & \longrightarrow & H^kU \cup V & \longrightarrow & H^kU \oplus H^kV & \longrightarrow & H^kU \cap V \\
\downarrow & & \downarrow & & \downarrow & & \downarrow & & \downarrow \\
H_{k-1}^\infty U^* \oplus H_{k-1}^\infty V^* & \longrightarrow & H_{k-1}^\infty U \cap V^* & \longrightarrow & H_k^\infty U \cup V^* & \longrightarrow & H_k^\infty U^* \oplus H_k^\infty V^* & \longrightarrow & H_k^\infty U \cap V^*.
\end{array}$$

An application of the five-lemma, together with the following exercise, finishes the proof. □

Exercise 24.14. Show that this diagram commutes.

These theorems justify the use of $\mathscr{C}^\infty$ techniques in studying the topology of a differentiable manifold. For example, they show that the De Rham groups depend only on the underlying topology of the manifold. Combining the isomorphisms of the two theorems, one has justified writing $\int_z \omega$ for z a continuous k-cycle and ω a closed $\mathscr{C}^\infty$ k-form on a manifold.

Problem 24.15. Show that for U open in $\mathbb{R}^n$, two classes τ_1 and τ_2 in $H_{n-1}U$ are equal if and only if $\int_{\tau_1} \omega = \int_{\tau_2} \omega$ for all closed $(n-1)$-forms ω on U.

Exercise 24.16. Let X be an n-manifold that can be covered by a finite number of open sets such that any intersection of them is diffeomorphic to a convex open set in $\mathbb{R}^n$. (It is a fact, proved by using a Riemannian metric and geodesics, that any compact manifold has such an open cover.) Show that each H_kX is a finitely generated abelian group, and that each H^kX and H^k_cX is a finite-dimensional vector space.

24c. Cohomology and Cohomology with Compact Supports

In higher dimensions, except in simple cases in the Poincaré lemmas, we have not yet used the higher-dimensional versions of wedging forms that we used on surfaces in Chapter 18. In general the wedge $\omega \wedge \mu$ of a k-form ω and an l-form μ is a $(k+l)$-form. This operation is linear in each factor, and satisfies the identities:

(i) $\mu \wedge \omega = (-1)^{k \cdot l} \omega \wedge \mu$; and
(ii) $d(\omega \wedge \mu) = d\omega \wedge \mu + (-1)^k \omega \wedge d\mu$.

Again, we assume these properties from advanced calculus. If either ω or μ has compact support, then $\omega \wedge \mu$ has compact support, since the support of the wedge product is contained in the intersection of the supports of the factors. It follows from (ii) that the wedge product of two closed forms is closed, and that, if one is closed and the other is exact, the wedge product is exact. From this it follows that the wedge product determine products on the cohomology groups

$$\wedge : H^kX \times H^lX \to H^{k+l}X$$

and

$$\wedge : H^k X \times H^l_c X \to H^{k+l}_c X,$$

each by the formula $[\omega] \times [\mu] \mapsto [\omega] \wedge [\mu] = [\omega \wedge \mu]$.

Exercise 24.17. Verify that these are well-defined bilinear mappings. Show that the first satisfies the formula $[\mu] \wedge [\omega] = (-1)^{k \cdot l}[\omega] \wedge [\mu]$. Prove that these products are associative where defined.

Now suppose X is oriented. As we saw in §22e, integrating over the manifold gives a mapping $H^n_c X \to \mathbb{R}$. So we have homomorphisms

$$H^k X \times H^{n-k}_c X \xrightarrow{\wedge} H^n_c X \to \mathbb{R}.$$

This determines linear maps $\mathscr{D}_X \colon H^k X \to \operatorname{Hom}(H^{n-k}_c X, \mathbb{R})$. Explicitly, $\mathscr{D}_X$ takes the class of a closed k-form ω to the homomorphism that takes the class of a closed $(n-k)$-form μ with compact support to the integral $\int_X \omega \wedge \mu$.

Theorem 24.18. *For any oriented manifold X the duality maps*

$$\mathscr{D}_X \colon H^k X \to \operatorname{Hom}(H^{n-k}_c X, \mathbb{R})$$

are isomorphisms.

Proof. The proof is almost identical to that for Theorems 24.12 and 24.13. This time, for (1), note that $H^0 U = \mathbb{R}$ and $H^n_c U \cong \mathbb{R}$ for U an open rectangle, and all other groups vanish, by the Poincaré lemmas; and since $1 \in H^0 U$ maps to the nonzero homomorphism that is integration over U, the map is an isomorphism. For (2), one again has maps from the Mayer–Vietoris sequence for H^k to the dual of the sequence for the H^{n-k}_c. This time the signs involved in the definition mean that the key square in the diagram only commutes up to sign, but that is good enough to apply the five-lemma, cf. Exercise 24.2. We leave the calculation of these signs as an exercise. □

This duality theorem has several corollaries that were not obvious before. For example, the simple fact that $H^0_c X = 0$ whenever X is a connected but not compact manifold implies the

Corollary 24.19. *If X is a connected, oriented, but noncompact n-manifold, then $H^n X = 0$.*

Corollary 24.20. *If X is a connected oriented n-manifold, then the map $H^n_c X \to \mathbb{R}$, $\omega \mapsto \int_X \omega$, is an isomorphism.*

Proof. Since $H^0X = \mathbb{R}$, it follows from the theorem that H^n_cX is one dimensional, and we have seen that the map $H^n_cX \to \mathbb{R}$ is not zero. □

Problem 24.21. Let X be a nonorientable connected n-manifold, and let $p: \tilde{X} \to X$ be the orientation covering of §16a. (a) Construct maps $p^*: H^k_cX \to H^k_c\tilde{X}$ and $p_*: H^k_c\tilde{X} \to H^k_cX$ so that $p_* \circ p^*\mu = 2 \cdot \mu$ and $p^* \circ p_*\omega = \omega + \tau_*\omega$, where $\tau: \tilde{X} \to \tilde{X}$ is the nontrivial deck transformation. (b) Deduce that p^* embeds H^k_cX as the subspace of $H^k_c\tilde{X}$ consisting of classes of the form $\omega + \tau_*\omega$. (c) Show that $H^n_cX = 0$. In particular, if X is compact, then $H^nX = 0$.

Corollary 24.22 (Poincaré Duality). *If X is a compact oriented n-manifold, then the pairing $H^kX \times H^{n-k}X \to \mathbb{R}$ is a perfect pairing, i.e., for any linear map $\varphi: H^{n-k}X \to \mathbb{R}$, there is a unique ω in H^kX such that $\varphi(\mu) = \int_X \omega \wedge \mu$ for all μ in $H^{n-k}X$.*

Problem 24.23. (a) Use this corollary to prove that H^kX is finite dimensional. (b) If $n = 2m$, with m odd, show that the dimension of H^mX is even, and deduce that the Euler characteristic

$$\sum_{k=0}^{n} (-1)^k \dim(H^kX)$$

must be even.

This puts strong restrictions on the homology and cohomology groups of a compact oriented n-manifold. For example, the dimension of H^kX must equal the dimension of $H^{n-k}X$. The skew-commutative algebra structure on the direct sum of the cohomology groups is also useful in many applications.

As we saw for Riemann surfaces these duality theorems can be used to define an *intersection number* $\langle\alpha, \beta\rangle$ for homology classes α in H_pX and β in $H_{n-p}X$, when X is an oriented n-manifold. As in that case, it is possible to do this directly and geometrically, by finding representative cycles that meet transversally, and counting the points of intersection with an appropriate sign. This takes quite a bit of work, however, and one can use duality to define the intersection number quickly: A class α in H_pX determines a linear map from H^pX to $\mathbb{R}$ by $\mu \mapsto \int_\alpha \mu$, and by Poincaré duality there is a unique class ω_α in $H^{n-p}X$ so that $\int_\alpha \mu = \int_X \omega_\alpha \wedge \mu$ for all μ in H^pX. By the same construction,

β in $H_{n-p}X$ determines ω_β in H^pX. So we can define

$$\langle \alpha, \beta \rangle = \int_X \omega_\alpha \wedge \omega_\beta .$$

This is a bilinear pairing, satisfying $\langle \beta, \alpha \rangle = (-1)^{p \cdot (n-p)} \langle \alpha, \beta \rangle$. The fact that $\langle \alpha, \beta \rangle$ is always an integer, however, is not so obvious from this definition, although it can often be verified directly by making constructions for representatives of ω_α and ω_β, as we did for surfaces in Chapter 18.

Exercise 24.24. Suppose X is oriented but not necessarily compact, and X has an open cover as in Exercise 24.16. Construct a homomorphism

$$H_pX \to H_c^{n-p}X, \qquad \alpha \mapsto \omega_\alpha ,$$

characterized by the equality $\int_\alpha \mu = \int_X \omega_\alpha \wedge \mu$ for all μ in H^pX.

Exercise 24.25. Suppose a topological space is a union of an increasing family of open subsets U_i, $U_1 \subset U_2 \subset \ldots$. Show that any element of H_kX is the image of an element of some H_kU_i, and that α_i in H_kU_i and α_j in H_kU_j determine the same element of H_kX if and only if there is some $m \geq \max(i,j)$ such that α_i and α_j have the same image in H_kU_m. This is expressed by saying that H_kX is the *direct limit* of the H_kU_i, and written

$$H_kX = \varinjlim H_kU_i .$$

Exercise 24.26. Suppose a manifold X is an increasing union of open subsets U_i, $U_1 \subset U_2 \subset \ldots$. (a) Use duality to deduce that giving a class η in H^kX is equivalent to giving a collection of classes η_i in H^kU_i for all i such that η_i restricts to η_j if $i > j$. This says that H^kX is the *inverse limit* of the H^kU_i, and is written

$$H^kX = \varprojlim H^kU_i .$$

(Note that this is not obvious from the definition of De Rham groups, even for open sets in $\mathbb{R}^n$.) (b) Show that H_c^kX is the direct limit of the $H_c^kU_i$:

$$H_c^kM = \varinjlim H_c^kU_i .$$

In fact, one can construct cohomology groups $H^k(X;\mathbb{Z})$ for any space X, which are finitely generated abelian groups for manifolds as in Exercise 24.16, and one can find an analogue of the wedge product

for these groups; after proving appropriate duality theorems, one has a construction of the intersection pairing whose values are integers. This could be a next chapter, if this book didn't end here. At least now we can give a quick definition of these groups, or of cohomology groups $H^k(X;G)$ with coefficients in any abelian group G, generalizing directly the discussion in §16c. Define a *k-cochain* to be an arbitrary function that assigns to every nondegenerate k-cube in X an element of G; these form a group $C^k(X;G)$. If c is a k-cochain, define the *coboundary* $\delta(c)$ of c to be the $(k+1)$-cochain defined by the formula $\delta(c)(\Gamma) = c(\partial\Gamma)$, where a cochain is extended linearly to be defined on all chains. Then $\delta \circ \delta = 0$, so one can define $H^k(X;G) = Z^k(X;G)/B^k(X;G)$, where $Z^k(X;G)$ is the group of k-cocycles (whose boundary is zero), and $B^k(X;G)$ is the group of k-coboundaries (of $(k-1)$-cochains).

Exercise 24.27. Prove that these groups satisfy the same properties as homology groups, but "dual." For example, maps $f\colon X \to Y$ determine (functorial) homomorphisms $f^*\colon H^kY \to H^kX$, homotopic maps determine the same maps on cohomology groups. State and prove the Mayer–Vietoris theorem for these groups. Construct homomorphisms from $H^k(X;G)$ to $\mathrm{Hom}(H_kX, G)$, and show that these are isomorphisms if $G = \mathbb{R}$.

Project 24.28. If G is an abelian group, and $\mathcal{U}$ is an open covering of a space X, define and study Čech groups $H^k(\mathcal{U};G)$ generalizing the groups $H^1(\mathcal{U};G)$ studied in Chapter 15.

24d. Simplicial Complexes

We have seen the usefulness of triangulating a surface. Many spaces that arise in nature, including many which are not manifolds, admit triangulations. When a space is triangulated, there is a much smaller chain complex that can be used to compute its homology. The general methods of §24a can be used to show that this complex computes the same homology as that using cubical chains.

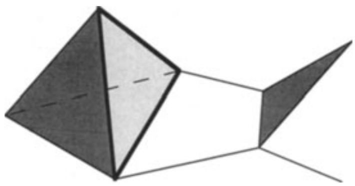
A (finite) *abstract simplicial complex* is a finite set V, called the vertices, and a collection K of subsets of V, called the (abstract) *simplices*, with the property that every subset of a simplex is a simplex. One usually assumes also that every set $\{v\}$ for v in V is a simplex, and one says that K is the simplicial complex. An *n-simplex* is a set σ in K with $n+1$ elements. A subset τ of a simplex σ is called a *face* of σ.

A set of $n+1$ points $P_0, \ldots, P_n$ in a vector space is called *affinely independent* if there is no relation $t_0P_0 + t_1P_1 + \ldots + t_nP_n = 0$ with $t_0, \ldots, t_n$ real numbers satisfying $t_0 + t_1 + \ldots + t_n = 0$ with not all $t_i = 0$. Equivalently, the vectors $P_1 - P_0, P_2 - P_0, \ldots, P_n - P_0$ are linearly independent. In this case the set of points

$$\{t_0P_0 + t_1P_1 + \ldots + t_nP_n \colon\ t_i \geq 0,\ t_0 + t_1 + \ldots + t_n = 1\}$$

is called the (geometric) *simplex* spanned by the points. It is homeomorphic to an n-dimensional disk.

The *realization* $|K|$ of an abstract simplicial complex K can be constructed by taking the vertices V to be the basis vectors for a vector space, and defining $|K|$ to be the union of the geometric simplices spanned by the abstract simplices in K. In practice one often takes the vertices in a smaller vector space, provided those in any simplex are affinely independent, and two geometric simplices are either disjoint or meet only along common faces.



We want to write down a chain complex for the simplicial complex K. This is simplest if K is *ordered*. This means that a partial ordering is given for the vertices, such that the vertices of each simplex are totally ordered. Each simplex σ then has a unique representation $\sigma = (v_0, \ldots, v_n)$ where the vertices of σ are listed in order. The chain complex C_*K of the ordered simplicial complex K is defined as follows: C_nK is the free abelian group on the n-simplices of K, and the boundary $\partial\colon C_nK \to C_{n-1}K$ is defined by

$$(24.29)\quad \partial((v_0, \ldots, v_n)) = \sum_{i=0}^{n} (-1)^i (v_0, \ldots, v_{i-1}, v_{i+1}, \ldots, v_n).$$

Exercise 24.30. Verify that the composite $\partial \circ \partial\colon C_nK \to C_{n-2}K$ is zero.

The nth homology group $H_n(C_*K)$ of this complex is denoted H_nK.

Exercise 24.31. Suppose K has a vertex v_0 with the property that for every simplex σ in K, the subset consisting of σ and v_0 is also in K;

denote this simplex by (v_0, σ). (Geometrically, $|K|$ is a cone with vertex v_0.) Assume that K is ordered so that v_0 comes before all other vertices. Define maps $H\colon C_nK \to C_{n+1}K$ by the formula

$$H(\sigma) = \begin{cases} (v_0, \sigma) & \text{if } v_0 \text{ is not a vertex of } \sigma, \\ 0 & \text{if } v_0 \text{ is a vertex of } \sigma. \end{cases}$$

Show that $\partial \circ H + H \circ \partial$ is the identity map on C_nK for $n > 0$. Deduce that $H_nK = 0$ for $n > 0$, and that $H_0K \cong \mathbb{Z}$.

A *subcomplex* L of a simplicial complex K is subset of the simplices in K such that whenever a simplex σ is in L, so are all its faces; then L is a simplicial complex, with its vertices a subset of the vertices of V. An ordering of K determines an ordering of L, and one has a canonical map $C_*L \to C_*K$, determining homomorphisms $H_nL \to H_nK$ on homology groups. If L_1 and L_2 are subcomplexes of K, the intersection $L_1 \cap L_2$ and union $L_1 \cup L_2$ are also subcomplexes. These maps determine an exact sequence of chain complexes

$$0 \to C_*(L_1 \cap L_2) \xrightarrow{-} C_*L_1 \oplus C_*L_2 \xrightarrow{+} C_*(L_1 \cup L_2) \to 0,$$

which determines a long exact Mayer–Vietoris sequence

$$\ldots \to H_{n+1}(L_1 \cup L_2) \to H_n(L_1 \cap L_2) \to H_nL_1 \oplus H_nL_2 \to H_n(L_1 \cup L_2) \to \ldots .$$

We want to compare the homology of K with the homology of its geometric realization $|K|$. For each ordered simplex $\sigma = (v_0, \ldots, v_n)$ we need to define a cubical n-chain $\Gamma_\sigma = \Gamma_{(v_0, \ldots, v_n)}$ in $|K|$. If $n = 0$, Γ_σ is the constant 0-chain at v_0. If $n = 1$, Γ_σ is the path from v_0 to v_1: $\Gamma_\sigma(t) = tv_1 + (1-t)v_0$. In general, define $\Gamma_\sigma\colon I^n \to |K|$ inductively by the formula

$$\Gamma_\sigma(t_1, \ldots, t_n) = t_nv_n + (1-t_n)\Gamma_{(v_0, \ldots, v_{n-1})}(t_1, \ldots, t_{n-1}).$$

Writing this out, we have

$$\text{(24.32)} \quad \Gamma_\sigma(t_1, \ldots, t_n) = \sum_{k=0}^{n} t_k(1-t_{k+1}) \cdot \ldots \cdot (1-t_n)v_k,$$

where, when $k = 0$, t_0 is set equal to 1.

Proposition 24.33. (*a*) *The map $\sigma \to \Gamma_\sigma$ determines a homomorphism $C_*K \to C_*|K|$ of chain complexes.* (*b*) *The induced homomorphisms $H_nK \to H_n|K|$ are isomorphisms.*

Proof. For (a), we must show that $\partial\Gamma_\sigma = \Sigma_{i=0}^n(-1)^i\Gamma_{(v_0,\ldots,\hat{v}_i,\ldots,v_n)}$, where the ^ denotes an omitted vertex. From the definition of $\partial\Gamma_\sigma$ as $\Sigma_{i=1}^k(-1)^i(\partial_i^0\Gamma_\sigma - \partial_i^1\Gamma_\sigma)$, this follows from the following three calculations, which are simple exercises, using (24.32):

(i) $\partial_1^1\Gamma_\sigma = \Gamma_{(v_1,\ldots,v_n)}$;
(ii) $\partial_i^1\Gamma_\sigma$ is a degenerate $(n-1)$-cube if $i > 1$; and
(iii) $\partial_i^0\Gamma_\sigma = \Gamma_{(v_0,\ldots,\hat{v}_i,\ldots,v_n)}$.

The proof of (b) will be by the (by now) familiar induction using Mayer–Vietoris, as follows. If L_1 and L_2 are subcomplexes of K, we have a commutative diagram

$$\begin{array}{ccccccccc} 0 & \longrightarrow & C_*(L_1\cap L_2) & \xrightarrow{-} & C_*L_1\oplus C_*L_2 & \xrightarrow{+} & C_*(L_1\cup L_2) & \longrightarrow & 0 \\ & & \downarrow & & \downarrow & & \downarrow & & \\ 0 & \longrightarrow & C_*(|L_1\cap L_2|) & \xrightarrow{-} & C_*|L_1|\oplus C_*|L_2| & \xrightarrow{+} & C_*(|L_1\cup L_2|) & \longrightarrow & 0. \end{array}$$

This gives a corresponding commutative diagram of long exact sequences, and the five-lemma shows that if (b) is true for L_1 and L_2 and $L_1\cap L_2$, then (b) is also true for $L_1\cup L_2$.

We can now prove (b) by induction on the number of simplices in the simplicial complex K. Take any vertex v of K. Let L_1 be the subcomplex consisting of all simplices of K that are contained in a simplex of K that contains v, and let L_2 be the subcomplex consisting of all simplices of K that do not contain v. Then (b) is known for L_1 by Exercise 24.31, and (b) is known for L_2 and $L_1\cap L_2$ by induction on the number of vertices. The preceding argument then shows that (b) holds for $K = L_1\cup L_2$. □

Corollary 24.34. *If K and L are simplicial complexes whose geometric realizations are homeomorphic, then $H_nK \cong H_nL$ for all n.*

Proof. This follows from the fact that a homeomorphism between spaces induces an isomorphism between their homology groups. □

In the early days of algebraic topology, the homology of a compact space X was defined by triangulating the space, i.e., finding a homeomorphism between some $|K|$ and X, and taking the homology H_*K. With this as the definition the assertion of the preceding corollary—that homology is a topological invariant of the space—was a serious problem.

The preceding discussion depended on a choice of ordering of the simplicial complex, which is how one would usually use the result in

calculations. The following exercise shows how this can be circumvented:

Exercise 24.35. For an abstract simplicial complex K, define C_nK to be the quotient of the free abelian group on the set of symbols $(v_0, \ldots, v_n)$, where $v_0, \ldots, v_n$ is an (ordered) set of vertices spanning an n-simplex of K, modulo the subgroup generated by relations

$$(v_0, \ldots, v_n) - \operatorname{sgn}(\tau)(v_{\tau(0)}, \ldots, v_{\tau(n)}),$$

for all permutations τ in the symmetric group ${}_{n+1}$, where $\operatorname{sgn}(\tau) = \pm 1$ is the sign of the permutation. Then C_nK is a free abelian group of rank equal to the number of n-simplices, but with basis elements only specified up to multiplication by ± 1. (a) Show that formula (24.29) determines a boundary map $\partial\colon C_nK \to C_{n-1}K$, with $\partial \circ \partial = 0$. (b) Given an ordered n-simplex $(v_0, \ldots, v_n)$, define $\Gamma'_{(v_0, \ldots, v_n)}$ to be $\Gamma_{(b_0, \ldots, b_n)}$, where b_k is the barycenter of the simplex spanned by the first $k+1$ vertices, i.e., $b_k = 1/(k+1)(v_0 + v_1 + \ldots + v_k)$. Define a map from C_*K to $C_*|K|$ by sending $(v_0, \ldots, v_n)$ to the sum $\Sigma \operatorname{sgn}(\tau)\Gamma'_{(v_{\tau(0)}, \ldots, v_{\tau(n)})}$, the sum over all τ in $\mathfrak{S}_{n+1}$. Show that this determines a homomorphism of chain complexes, and show that the resulting map in homology is an isomorphism. (c) Show that an ordering of K determines an isomorphism of the complex defined earlier with the complex defined in this exercise.

Problem 24.36. If c_i is the number of i-simplices in K, show that the Euler characteristic is the alternating sum of the numbers of simplices:

$$\Sigma(-1)^i c_i = \Sigma(-1)^i \dim(H_i(|K|)),$$

generalizing what we have seen for surfaces.

Problem 24.37. (a) If $\mathcal{U} = \{U_v, v \in V\}$ is a finite collection of open sets whose union is a space X, define a simplicial complex, called the *nerve* of $\mathcal{U}$ and denoted $N(\mathcal{U})$, by taking V to be the vertices, and defining the simplices to be the subsets S such that the intersection of the U_v for v in S is nonempty. Verify that $N(\mathcal{U})$ is a simplicial complex.

(b) If K is any simplicial complex, and v is a vertex in K, define an open set $\operatorname{St}(v)$ in $|K|$, called the *star* of v, to be the union the "interiors" of the simplices that contain v, i.e., $\operatorname{St}(v)$ is the complement in $|K|$ of the union of those $|\sigma|$ for which σ does not contain v. Show that the open sets $\{\operatorname{St}(v), v \in V\}$ form an open covering of $|K|$, and that the nerve of this covering is the same as K.

(c) Suppose $\mathcal{U}$ is an open covering of X as in (a), with the property that for all $v_0, \ldots, v_r$ in V, $U_{v_0} \cap \ldots \cap U_{v_r}$ is connected and $H_k(U_{v_0} \cap \ldots \cap U_{v_r}) = 0$ for all $k > 0$. Construct a homomorphism of chain complexes from $C_*(N(\mathcal{U}))$ to C_*X, and show that it determines an isomorphism from $H_k(N(\mathcal{U}))$ to H_kX for all k.

APPENDICES

These appendices collect some facts used in the text. The beginnings of Appendices A, B, and C state definitions and basic results from point set topology, calculus, and algebra that should be reasonably familiar, together with proofs of a few basic results that may be slightly less so. Each of these appendices ends with some more technical results that may be consulted as the need arises. Appendix D contains two technical lemmas about vector fields in the plane, as well as some basic definitions about coordinate charts and differential forms on surfaces. Appendix E contains a proof of Borsuk's general theorem on antipodal maps that was stated in Chapter 23.

Conventions and Notation

A *closed rectangle* in $\mathbb{R}^2$ has sides parallel to the axes, so is a subset of the form $[a,b] \times [c,d]$, with $a < b$ and $c < d$. An *open rectangle* is a product of two open intervals, usually finite, but we occasionally allow them to be infinite.

The *unit interval* I is $[0,1] = \{x \in \mathbb{R}: 0 \leq x \leq 1\}$.
The *n-dimensional disk* D^n is

$$D^n = \{(x_1, \ldots, x_n) \in \mathbb{R}^n: x_1^2 + \ldots + x_n^2 \leq 1\}.$$

The *n-sphere* S^n is

$$S^n = \{(x_1, \ldots, x_{n+1}) \in \mathbb{R}^{n+1}: x_1^2 + \ldots + x_{n+1}^2 = 1\}.$$

The origin $(0, 0, \ldots, 0)$ in $\mathbb{R}^n$ is often denoted simply by 0.

APPENDIX A

Point Set Topology

A1. Some Basic Notions in Topology

A *topology* on a set X is a collection of subsets, called the open sets, including X itself and the empty set, such that any union of open sets is open, and any finite intersection of open sets is open. A *topological space* is a set X together with a topology. A collection of open sets is a *basis* for the open sets if any open set is a union of sets in the basis. For example, if X is a metric space, the open balls $B_\varepsilon(x) = \{y \in X: \text{distance}(y, x) < \varepsilon\}$ form a basis for a topology on X. In particular, Euclidean space $\mathbb{R}^n$ with its usual distance function is a topological space. A *neighborhood* of a point in a topological space is an open set containing the point—or, occasionally, any set containing such an open set.

Any subset Y of a topological space is a topological space with the induced topology: the open sets are those of the form $U \cap Y$, for U open in X. Such Y is called a *topological subspace* of X. In particular, any subset of $\mathbb{R}^n$ is a topological space. A subset Y is *closed* if its complement is open. A map $f: X \rightarrow Y$ from one topological space to another is *continuous* if $f^{-1}(U)$ is open in X for every open set U in Y. A bijection $f: X \rightarrow Y$ is a *homeomorphism* if f and f^{-1} are continuous.

A topological space X is *Hausdorff* if, for any two distinct points in X, there are disjoint open sets, one containing one of the points, the other containing the other. Any metric space is Hausdorff. Although we seldom need to assume spaces are Hausdorff, the reader will lose little by assuming that all spaces occurring in the book are Hausdorff.

A subset K of a space X is called *compact*, if, for any collection of open

sets $\{U_\alpha: \alpha \in \mathscr{A}\}$ such that K is contained in the union of the U_α, there is a finite subset $\{\alpha(1), \ldots, \alpha(m)\}$ of $\mathscr{A}$ so that K is contained in the union $U_{\alpha(1)} \cup \ldots \cup U_{\alpha(m)}$. The following are some basic facts about compact spaces:

(A.1) If $f: X \to Y$ is continuous, and K is a compact subset of X, then $f(K)$ is a compact subset of Y.
(A.2) A compact subset of a Hausdorff space is closed.
(A.3) If $f: X \to Y$ is continuous and bijective, and X is compact and Y is Hausdorff, then f is a homeomorphism.
(A.4) A subset K of $\mathbb{R}^n$ is compact if and only if it is closed and bounded.

Exercise A.5. If K and L are disjoint compact subsets in a Hausdorff space X, show that there are disjoint open sets in X, one containing K, the other containing L.

Exercise A.6. If K is compact, and, for each positive integer n, A_n is a nonempty subset of K, show that there is a limit point, i.e., a point P in K such that every neighborhood of P meets A_n for an infinite number of integers n.

Exercise A.7. (a) Show that a rectangle $[a, b] \times [c, d]$ is homeomorphic to the closed unit disk $\{(x, y): x^2 + y^2 \leq 1\}$. (b) Show that $\mathbb{R}^2$ is homeomorphic to the open unit disk $\{(x, y): x^2 + y^2 < 1\}$.

A subset X of $\mathbb{R}^n$ is *convex* if, for any points P and Q in X, the line segment $\{t \cdot P + (1 - t) \cdot Q: 0 \leq t \leq 1\}$ from P to Q is contained in X.

Problem A.8. Show that any compact, convex subset of $\mathbb{R}^n$ that contains a nonempty open set is homeomorphic to the closed unit disk

$$D^n = \{(x_1, \ldots, x_n): x_1{}^2 + \ldots + x_n{}^2 \leq 1\}.$$

If ∂K is the *boundary* of K, i.e., ∂K is the set of points of K such that every neighborhood contains points inside and outside K, show that there is a homeomorphism from K to D^n that maps ∂K homeomorphically onto the boundary $S^{n-1} = \partial D^n$.

If X and Y are topological spaces, the Cartesian products $U \times V$ of open sets U in X and V in Y form a basis for a topology in the Cartesian product $X \times Y$, called the *product* topology.

If X and Y are topological spaces, the *disjoint union* $X \coprod Y$ is a topological space. A set in the disjoint union is open when it is the union of an open set in X and an open set in Y. More generally, if $\{X_\alpha: \alpha \in \mathscr{A}\}$ is any collection of topological spaces, the disjoint union $\coprod X_\alpha$ is a topological space, with open sets disjoint unions of open sets in each X_α.

Any set X can be made into a topological space with the *discrete topology*, in which every subset of X is open. Equivalently, all points are open.

The *interior* of a subset A of a topological space, denoted Int(A), is the set of points that have a neighborhood contained in A. The *closure* of a subset A, denoted $\overline{A}$, is the intersection of all closed sets containing A.

A2. Connected Components

A topological space X is *connected* if it cannot be written as a union of two nonempty disjoint sets, each of which is open in X.

Exercise A.9. Show that each of the following is equivalent to X being connected: (i) X has no nonempty proper subset that is both open and closed; (ii) X cannot be written as the union of two nonempty disjoint closed subsets; and (iii) there is no continuous mapping from X onto the discrete space $\{0, 1\}$.

The following are basic facts about connected spaces:

(A.10) If $f: X \to Y$ is a continuous, surjective mapping, and X is connected, then Y is connected.
(A.11) If X is a subspace of a space Y, and X is connected, then the closure $\overline{X}$ of X in Y is also connected.
(A.12) If X is a union of a family of subspaces X_α, each of which is connected, and each pair of which have nonempty intersection, then X is connected.
(A.13) The connected subsets of $\mathbb{R}$ are the intervals.

A *connected component* of X is a connected subset that is not contained in any larger connected subset. Each connected component is closed in X. Any two connected components of X are disjoint. The union of all connected subsets of X containing a point x is a connected component, called the *connected component of x in X*. The space X is a disjoint union of its connected components.

Exercise A.14. Let $X \subset \mathbb{R}^2$ be the union of the points $(0, 0)$, $(0, 1)$, and the lines $\{1/n\} \times [0, 1]$, $n = 1, 2, \ldots$. Show that these are the connected components of X, but whenever X is written as the union of two open and closed subsets, the points $(0, 0)$ and $(0, 1)$ belong to the same subset.

A space is called *locally connected* if for every neighborhood V of every point x, there is a connected open neighborhood U of x that is contained in V.

Exercise A.15. If X is locally connected, show that all the connected components of X are open in X.

A space X is *path-connected* if, for any two points x and y in X, there is a continuous mapping γ from an interval $[a, b]$ to X that maps a to x and b to y. A space is *locally path-connected* if every neighborhood of every point contains a path-connected neighborhood of the point.

Exercise A.16. If X is locally path-connected, show that all the connected components of X are path-connected and open in X.

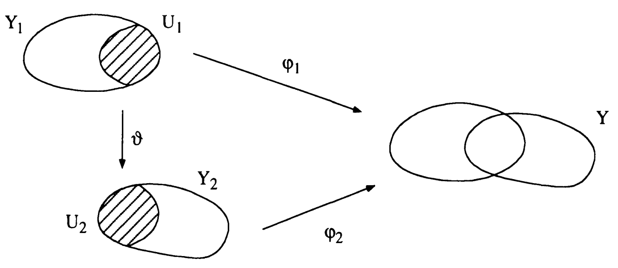
In particular, for U open in the plane or $\mathbb{R}^n$, the connected components of U are open and path-connected.

A3. Patching

If a topological space X is the union of two sets A and B, both open or both closed, and $f\colon A \to Y$ and $g\colon B \to Y$ are continuous mappings from A and B to a space Y, such that f and g agree on $A \cap B$, then there is a unique continuous mapping h from X to Y that agrees with f on A and with g on B.

If Y is a topological space, and R is any equivalence relation on Y, the set Y/R of equivalence classes is given the *quotient topology*: a set U is open in Y/R exactly when its inverse image in Y is open. If $f\colon Y \to Z$ is a continuous mapping that maps all points in each equivalence class to the same point, then f determines a continuous mapping $\bar{f}\colon Y/R \to Z$ so that the composite $Y \to Y/R \to Z$ is f.

Suppose Y_1 and Y_2 are two spaces, with open subsets U_1 of Y_1 and U_2 of Y_2, and a homeomorphism $\vartheta\colon U_1 \to U_2$ is given between them. Then one can *patch* (or *glue*, or *clutch*) the spaces Y_1 and Y_2 together, to form a space Y. There will be maps $\varphi_1\colon Y_1 \to Y$ and $\varphi_2\colon Y_2 \to Y$; Y will be the union of the open subsets $\varphi_1(Y_1)$ and $\varphi_2(Y_2)$, each φ_i will map Y_i homeomorphically onto $\varphi_1(Y_i)$, with $\varphi_1(U_1) = \varphi_2(U_2)$, and ϑ will be the composite $\varphi_2^{-1} \circ \varphi_1$ on U_1.



One can construct Y as the quotient space $Y_1 \coprod Y_2/R$, where R is the equivalence relation consisting of pairs $(u_1, \vartheta(u_1))$ for u_1 in U_1, and of course the symmetric pairs $(\vartheta(u_1), u_1)$, together with all pairs (y_1, y_1) for y_1 in Y_1 and all pairs (y_2, y_2) for y_2 in Y_2.

More generally, suppose we have a collection Y_α of spaces, for α in an index set $\mathcal{A}$, and, for each α and β in $\mathcal{A}$, we have an open subset $U_{\alpha\beta}$ of Y_α, and a homeomorphism

$$\vartheta_{\beta\alpha}\colon U_{\alpha\beta} \to U_{\beta\alpha}.$$

These should satisfy the conditions:

(1) $U_{\alpha\alpha} = Y_\alpha$, and $\vartheta_{\alpha\alpha}$ is the identity on Y_α; and
(2) for any α, β, and γ in $\mathcal{A}$, $\vartheta_{\beta\alpha}(U_{\alpha\beta} \cap U_{\alpha\gamma}) \subset U_{\beta\gamma}$ and

$$\vartheta_{\gamma\alpha} = \vartheta_{\gamma\beta} \circ \vartheta_{\beta\alpha} \qquad \text{on} \quad U_{\alpha\beta} \cap U_{\alpha\gamma}.$$

In particular, $\vartheta_{\alpha\beta} \circ \vartheta_{\beta\alpha}$ is the identity on $U_{\alpha\beta}$. Set

$$Y = \coprod_{\alpha \in \mathcal{A}} Y_\alpha / R,$$

where R is the equivalence relation: y_α in Y_α is equivalent to y_β in Y_β if and only if $y_\alpha \in U_{\alpha\beta}$, $y_\beta \in U_{\beta\alpha}$, and $\vartheta_{\beta\alpha}(y_\alpha) = y_\beta$.

Let φ_α be the map from Y_α to Y that takes a point to its equivalence class. Give Y the quotient topology, which means that a set U in Y is open if and only if each $\varphi_\alpha^{-1}(U)$ is open in Y_α.

Lemma A.17. (1) *Each $\varphi_\alpha(Y_\alpha)$ is open in Y;*
(2) *φ_α is a homeomorphism of Y_α onto $\varphi_\alpha(Y_\alpha)$;*
(3) *Y is the union of the sets $\varphi_\alpha(Y_\alpha)$;*
(4) *$\varphi_\alpha(U_{\alpha\beta}) = \varphi_\beta(U_{\beta\alpha})$; and*
(5) *on $U_{\alpha\beta}$, $\vartheta_{\beta\alpha} = \varphi_\beta^{-1} \circ \varphi_\alpha$.*

Proof. The fact that φ_α is one-to-one onto its image, and the assertions (3)–(5), are set-theoretic verifications, and left to the reader. The topology on Y is defined to make each φ_α continuous. To prove (1) and (2), it suffices to verify that if U is open in some Y_α, then $\varphi_\alpha(U)$ is open in Y, i.e., that, for all β, $\varphi_\beta^{-1}(\varphi_\alpha(U))$ is open in Y_β. But $\varphi_\beta^{-1}(\varphi_\alpha(U)) = \vartheta_{\beta\alpha}(U \cap U_{\alpha\beta})$, which is open since $U \cap U_{\alpha\beta}$ is open in $U_{\alpha\beta}$ and $\vartheta_{\beta\alpha}$ is a homeomorphism. □

Exercise A.18. Make a similar construction if each $U_{\alpha\beta}$ is a closed subset of Y_α.

A4. Lebesgue Lemma

We make frequent use of the following lemma:

Lemma A.19. (Lebesgue Lemma). *Given any covering of a compact metric space K by open sets, there is an $\varepsilon > 0$ such that any subset of K of diameter less than ε is contained in some open set in the covering.*

Proof. If not, there is for every integer n a subset A_n of K with diameter less than $1/n$ and not contained in any open set of the covering. From the fact that K is compact it follows that there is a limit point P, see Exercise A.6. Let U be an open set of the covering that contains P, and take $r > 0$ so all points within distance r of P are contained in U. There must be (infinitely many) n with $1/n < r/2$ such that A_n meets the open ball of radius $r/2$ around P. But such A_n must be contained in U, a contradiction. □

The following lemma will be used in Appendix B to construct a partition of unity:

Lemma A.20. *If U is an open set in $\mathbb{R}^n$, there is a sequence of compact subsets $K_1, K_2, \ldots$, whose union is U, and so that*

$$K_1 \subset \text{Int}(K_2) \subset K_2 \subset \text{Int}(K_3) \subset \ldots \subset K_n \subset \text{Int}(K_{n+1}) \subset \ldots.$$

Proof. Start with any countable sequence of open sets U_i that cover U such that the closure $\overline{U}_i$ is compact and contained in U; for example, one can take the U_i to be balls at centers with rational coordinates and rational radii. Take $K_1 = \overline{U}_1$. Then take $K_2 = \overline{U}_1 \cup \ldots \cup \overline{U}_p$, where p is minimal such that K_1 is contained in $U_1 \cup \ldots \cup U_p$, and so on: if $K_m = \overline{U}_1 \cup \ldots \cup \overline{U}_s$, take $K_{m+1} = \overline{U}_1 \cup \ldots \cup \overline{U}_t$ where t is minimal so that K_m is contained in $U_1 \cup \ldots \cup U_t$. □

Remark A.21. The lemma is true, with the same proof, when U is replaced by any manifold whose topology has a countable basis of open sets.

APPENDIX B

Analysis

B1. Results from Plane Calculus

We list the basic results from calculus that were used in Chapters 1 and 2. As in those chapters, for simplicity, differentiable functions on a closed interval or rectangle are assumed to have differentiable extensions to some open neighborhood. Integrals of continuous functions on a closed interval, or a closed rectangle, are defined as limits of Riemann sums. The next five basic facts from calculus are stated for easy reference, in the form we need. Consult your favorite calculus book for proofs.

(B.1) Fundamental Theorem of Calculus. *If a continuous function f is the derivative of a function F on an interval* $[a, b]$, *then*

$$\int_a^b f(x)\,dx = F(b) - F(a).$$

(B.2) Mean Value Theorem. *If f is continuous on an interval* $[a, b]$, *there is an* x^* *with* $a < x^* < b$ *such that*

$$\frac{1}{b-a}\int_a^b f(x)\,dx = f(x^*).$$

(B.3) Chain Rule. *If* $\gamma(t) = (x(t), y(t))$, $a \le t \le b$ *is a differentiable path on an interval* $[a, b]$, *and f is a differentiable function on a neighborhood of*

$\gamma([a,b])$, then $f\circ\gamma$ is differentiable on $[a,b]$, and

$$\frac{d}{dt}(f(\gamma(t))) = \frac{\partial f}{\partial x}(x(t),y(t))\frac{dx}{dt}+\frac{\partial f}{\partial y}(x(t),y(t))\frac{dy}{dt}.$$

(B.4) Equality of Mixed Partial Derivatives. *If f is a $\mathscr{C}^\infty$ function on an open set in the plane, then*

$$\frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y}\right) = \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right).$$

(B.5) Fubini's Theorem. *If f is a continuous function on a rectangle $R=[a,b]\times[c,d]$, then*

$$\iint\limits_R f(x,y)\,dx\,dy = \int_a^b\left[\int_c^d f(x,y)\,dy\right]dx = \int_c^d\left[\int_a^b f(x,y)\,dx\right]dy.$$

Proposition B.6 (Green's Theorem for a Rectangle). *If p and q are continuously differentiable functions on a rectangle $R=[a,b]\times[c,d]$, then*

$$\iint\limits_R\left(\frac{\partial q}{\partial x}-\frac{\partial p}{\partial y}\right)dx\,dy = \int_a^b p(x,c)\,dx+\int_c^d q(b,y)\,dy$$
$$-\int_a^b p(x,d)\,dx-\int_c^d q(a,y)\,dy.$$

Proof. By Fubini's theorem and the fundamental theorem of calculus,

$$\iint\limits_R\frac{\partial q}{\partial x}dx\,dy = \int_c^d\left[\int_a^b\frac{\partial q}{\partial x}dx\right]dy = \int_c^d[q(b,y)-q(a,y)]\,dy;$$

$$\iint\limits_R\frac{\partial p}{\partial y}dx\,dy = \int_a^b\left[\int_c^d\frac{\partial q}{\partial y}dy\right]dx = \int_a^b[p(x,d)-p(x,c)]\,dx.$$

Green's theorem results by subtracting these two equations. □

Writing $\omega=p(x,y)\,dx+q(x,y)\,dy$, this says that

$$\iint\limits_R d\omega = \int_{\gamma_1}\omega+\int_{\gamma_2}\omega-\int_{\gamma_3}\omega-\int_{\gamma_4}\omega,$$

where γ_1, γ_2, γ_3, and γ_4 are the four sides of the rectangles, as in Chapter 1.

Corollary B.7. *If $d\omega=0$, then*

$$\int_{\gamma_1}\omega+\int_{\gamma_2}\omega = \int_{\gamma_3}\omega+\int_{\gamma_4}\omega.$$

Exercise B.8. If f is continuous on $[a,b]$, and $|f(t)| \leq M$ on $[a,b]$, show that $|\int_a^b f(t)\,dt| \leq M \cdot (b-a)$.

The following result will be used in Appendix D:

Lemma B.9. *If f is a $\mathscr{C}^\infty$ function in a neighborhood of $P = (a,b)$ in $\mathbb{R}^2$, with $f(P) = 0$, then there are $\mathscr{C}^\infty$ functions f_1 and f_2 so that*

$$f(x,y) = (x-a)f_1(x,y) + (y-b)f_2(x,y)$$

in a neighborhood of P.

Proof. We may assume $P = (0,0)$. By the fundamental theorem of calculus and the chain rule,

$$\begin{aligned} f(x,y) &= \int_0^1 \frac{\partial}{\partial t}(f(tx,ty))\,dt \\ &= x\int_0^1 \frac{\partial f}{\partial x}(tx,ty)\,dt + y\int_0^1 \frac{\partial f}{\partial y}(tx,ty)\,dt. \end{aligned}$$

The functions

$$f_1(x,y) = \int_0^1 \frac{\partial f}{\partial x}(tx,ty)\,dt \quad \text{and} \quad f_2(x,y) = \int_0^1 \frac{\partial f}{\partial y}(tx,ty)\,dt$$

are the required $\mathscr{C}^\infty$ functions. □

B2. Partition of Unity

For construction of the Mayer–Vietoris sequence for open sets in the plane, we need the following result:

Proposition B.10 (Partition of Unity). *Suppose an open set U in $\mathbb{R}^n$ is the union of a sequence $U_1, U_2, \ldots$ of open sets with the property that each point is contained in U_i for only finitely many i. Then there is a sequence of nonnegative $\mathscr{C}^\infty$ functions φ_i on U such that the closure (in U) of the support of φ_i is contained in U_i, and $\Sigma_{i=1}^\infty \varphi_i \equiv 1$ on U.*

(We will construct these functions so that only finitely many φ_i are nonzero in a neighborhood of any point of U, so the sum is a well-defined $\mathscr{C}^\infty$ function.)

Proof. There are several steps.

Step 1. There is a $\mathscr{C}^\infty$ function f on $\mathbb{R}$ that is zero on the negative half line and positive on the positive half line. Such a function is

$$f(x) = \begin{cases} \exp\left(-\dfrac{1}{x}\right) & \text{if } x > 0, \\ 0 & \text{if } x \leq 0. \end{cases}$$

Exercise B.11. Verify that this is infinitely differentiable by showing that any derivative of $\exp(-1/x)$ has the form $(P(x)/x^m)\exp(-1/x)$ for some polynomial P and some exponent m.

Step 2. Given a bounded rectangle $(a_1, b_1) \times \ldots \times (a_n, b_n)$, there is a $\mathscr{C}^\infty$ function h in $\mathbb{R}^n$ that is positive on the rectangle, and zero outside the rectangle. In fact, the function $g(x) = f(x) \cdot f(1-x)$ is positive on $(0, 1)$ and zero outside this interval, and this leads to the required function

$$h(x_1, \ldots, x_n) = \prod_{i=1}^{n} g\left(\frac{x_i - a_i}{b_i - a_i}\right).$$

Step 3. If A is a compact subset of U, there are $\mathscr{C}^\infty$ functions σ_i so that the closure of the support of σ_i is contained in U_i, and the sum $\Sigma_{i=1}^{\infty} \sigma_i$ is everywhere positive on A. To construct them cover A by a finite number of rectangles R_α such that the closure of each R_α is contained in some U_i, and use Step 2 to construct h_α that are positive on R_α and zero outside. Take σ_1 to be the sum of those h_α such that $\bar{R}_\alpha$ is contained in U_1, let σ_2 be the sum of those among the others such that the closure of the support is contained in U_2, and continue in this way until all h_α are used; set the other σ_i equal to zero.

Step 4. To complete the proof, write U as an increasing union of compact sets $K_1 \subset K_2 \subset \ldots$ as in Lemma A.20. Let $A_j = K_j \setminus \mathrm{Int}(K_{j-1})$ (where we set $K_j = \varnothing$ for $j \leq 0$). Let $W_j = \mathrm{Int}(K_{j+1}) \setminus K_{j-2}$. Note that A_j is a compact subset of the open set W_j. Apply Step 3 to each compact set $A_j \subset W_j$ with its open covering $\{U_i \cap W_j\}$, obtaining functions σ_{ij} so that the closure of the support of σ_{ij} is contained in $U_i \cap W_j$ and with $\Sigma_{i=1}^{\infty} \sigma_{ij}$ everywhere positive on A_j. Define ψ_i to be the sum $\Sigma_{j=1}^{\infty} \sigma_{ij}$. Only finitely many (at most three) terms in this sum are nonzero in some neighborhood of any point, so ψ_i is a $\mathscr{C}^\infty$ function the closure of whose support is contained in U_i. The sum $\psi = \Sigma_{i=1}^{\infty} \psi_i$ is similarly a $\mathscr{C}^\infty$ function, and ψ is positive on each A_j, so it is positive on all of U. Now

$$\varphi_i = \frac{\psi_i}{\psi}$$

satisfies all the required conditions. □

Exercise B.12 (Partition of Unity for Arbitrary Coverings). Suppose

$\mathcal{U} = \{U_\alpha : \alpha \in \mathcal{A}\}$ is an arbitrary open covering of an open set U in $\mathbb{R}^n$. Show that there is a sequence of nonnegative $\mathscr{C}^\infty$ functions $\varphi_1, \varphi_2, \ldots$ on U such that: (i) the closure (in U) of the support of φ_i is contained in some open set $U_{\alpha(i)}$; (ii) for each P in U there is a neighborhood of P such that only finitely many φ_i are nonzero on the neighborhood; and (iii) $\sum_{i=1}^{\infty} \varphi_i \equiv 1$.

Exercise B.13. Extend these results on partitions of unity to the case where U is any manifold that has a countable basis of open sets.

Exercise B.14. For $0 < r_1 < r_2$ construct a $\mathscr{C}^\infty$ function ψ on the plane that vanishes on the disk of radius r_1 centered at the origin, and is identically 1 outside the disk of radius r_2 centered at the origin, and takes values in the interval $(0, 1)$ between the two circles.

APPENDIX C

Algebra

C1. Linear Algebra

In this book, unless otherwise stated, vector spaces are real vector spaces. The vector space $\mathbb{R}^n$ consists of n-tuples $(x_1, \ldots, x_n)$ of real numbers, with coordinatewise addition and multiplication by scalars. A set of elements $\{e_\alpha\}$ is a *basis* for a vector space if every element in the space has a unique expression in the form $\Sigma x_\alpha e_\alpha$, for some real numbers x_α with only finitely many x_α nonzero. A vector space is *finite dimensional* if it has a finite basis. The number of elements in a basis is independent of choice of basis, and is the *dimension* of the space; the dimension of V is denoted $\dim(V)$. Choosing a basis $e_1, \ldots, e_n$ for V sets up an isomorphism of V with $\mathbb{R}^n$, with the vector $x_1e_1 + \ldots + x_ne_n$ in V corresponding to $(x_1, \ldots, x_n)$ in $\mathbb{R}^n$. The *standard basis* of $\mathbb{R}^n$ is the basis $\{e_i\}$ where e_i has a 1 for its ith coordinate, and zeros for the other coordinates.

If $L\colon V \to W$ is a linear mapping, the *kernel* $\mathrm{Ker}(L)$ is the subspace of V consisting of vectors mapped to zero, and the *image* $\mathrm{Im}(L)$ is the subspace of W consisting of vectors that can be written $L(v)$ for some v in V.

The *rank-nullity theorem* asserts that if $L\colon V \to W$ is a linear mapping of finite-dimensional vector spaces,

$$\dim(\mathrm{Ker}(L)) + \dim(\mathrm{Im}(L)) = \dim(V).$$

If W is a subspace of a vector space V, the *quotient space* V/W is defined to be the set of equivalence classes of elements of V, two vectors in V being equivalent if their difference is in W. This set V/W has a natural structure of a vector space, so that the mapping from V to V/W that takes a vector

to its equivalence class is a linear mapping of vector spaces. The kernel of this mapping from V to V/W is W.

Conversely, if $V \to U$ is a surjective linear mapping of vector spaces, and W is the kernel, this determines an isomorphism of V/W with U.

Suppose $L: V \to V'$ is a linear mapping of vector spaces, and W is a subspace of V, and W' a subspace of V'. If $L(W)$ is contained in W', then L determines a linear mapping.

$$V/W \to V'/W'$$

of quotient spaces, which takes the class of v in V to the class of $L(v)$ in V'.

If V and W are vector spaces, the *direct sum* $V \oplus W$ can be defined as the set of pairs (v, w), with v in V and w in W, with addition defined by $(v, w) + (v', w') = (v + v', w + w')$, and multiplication by scalars by $r \cdot (v, w) = (r \cdot v, r \cdot w)$. For example, $\mathbb{R}^n$ is the direct sum of n copies of $\mathbb{R}$. More generally, given any collection V_α of vector spaces, for α in some index set $\mathcal{A}$, an element of direct sum $\oplus V_\alpha$ is determined by specifying a vector v_α in V_α for each α in $\mathcal{A}$, with the added condition that v_α can be nonzero for only finitely many α. Addition and multiplication by scalars are defined component by component, as for two factors. The same definition, but without the restriction that only finitely many are nonzero, defines the *direct product*, denoted ΠV_α.

For vectors $u = (x_1, \ldots, x_n)$ and $v = (y_1, \ldots, y_n)$ in $\mathbb{R}^n$, the *dot product* is the number $u \cdot v = x_1 y_1 + \ldots + x_n y_n$. The *length* of u is $\|u\| = \sqrt{u \cdot u}$. The *projection* of u on v is the vector tv, where $t = (u \cdot v)/(v \cdot v)$; the length of this projection is $|u \cdot v|/\|v\|$.

An m by n matrix $A = (a_{i,j})$ determines a linear mapping $L: \mathbb{R}^n \to \mathbb{R}^m$ that takes e_j to $L(e_j) = \Sigma_{i=1}^m a_{i,j} e_i$. Every linear mapping from $\mathbb{R}^n$ to $\mathbb{R}^m$ arises from a unique such matrix. If $M: \mathbb{R}^m \to \mathbb{R}^l$ corresponds to an l by m matrix $B = (a_{j,k})$, the composite $M \circ L: \mathbb{R}^n \to \mathbb{R}^l$ corresponds to the *product* matrix $B \cdot A$, where the (i, j) entry of $B \cdot A$ is $\Sigma_{k=1}^m b_{i,k} a_{k,j}$. We need this mainly for (2×2) matrices, where a matrix $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is a matrix, the linear mapping corresponding to A takes a vector $v = (x, y)$ to the vector $(ax + by, cx + dy)$. The determinant of A, denoted $\det(A)$, is $ad - bc$. If the determinate is nonzero, A is invertible, with inverse

$$A^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

The invertible (2×2)-matrices form a group, denoted $\mathrm{GL}_2 \mathbb{R}$. This group has a topology, determined by its embedding as an open subset of $\mathbb{R}^4$: the complement of the set of (a, b, c, d) with $ad - bc = 0$.

Exercise C.1. Show that the product $\mathrm{GL}_2 \mathbb{R} \times \mathrm{GL}_2 \mathbb{R} \to \mathrm{GL}_2 \mathbb{R}$, $A \times B \mapsto A \cdot B$, and the inverse map $\mathrm{GL}_2 \mathbb{R} \to \mathrm{GL}_2 \mathbb{R}$, $A \mapsto A^{-1}$, are continuous mappings.

Lemma C.2. *If det(A)* > 0, *there is a path* γ: $[a,b] \to GL_2\mathbb{R}$ *such that* $\gamma(a) = A$ *and* $\gamma(b) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$. *If det(A)* < 0, *there is a path* γ: $[a,b] \to GL_2\mathbb{R}$ *such that* $\gamma(a) = A$ *and* $\gamma(b) = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$.

Proof. We will find a sequence of paths, each taking the matrix to a simpler one. For example, by multiplying a column of a matrix by t, with t varying in $[\alpha, \beta]$, for α and β positive, we can find a path changing the lengths of the column. In particular, we can assume the first column of A is a unit vector, so it can be written in the form $(\cos(\vartheta), \sin(\vartheta))$ for some ϑ in $[0, 2\pi]$. Then the path

$$\gamma(t) = \begin{bmatrix} \cos(t) & \sin(t) \\ -\sin(t) & \cos(t) \end{bmatrix} \cdot A, \qquad 0 \le t \le \vartheta,$$

takes A to a matrix whose first column is $(1,0)$. Then one can gradually project the second column on the line perpendicular to the first, via the path

$$\gamma(t) = \begin{bmatrix} 1 & (1-t)b \\ 0 & d \end{bmatrix}, \qquad 0 \le t \le 1,$$

to get to a matrix where the second column is $(0, d)$. Changing the length of the second column as at the beginning, we can get it either to $(0, 1)$ or to $(0, -1)$, as asserted. □

Problem C.3. Generalize to $GL_n\mathbb{R}$, showing that $GL_n\mathbb{R}$ has two connected components for all $n \geq 1$.

C2. Groups; Free Abelian Groups

A set of elements in a group G *generates* G if every element in G can be written as a (finite) product of elements in the set and inverses of elements in the set.

If H is a subgroup of a group G, a *left coset* is a subset of G of the form $g \cdot H = \{g \cdot h \colon h \in H\}$. The group G is a disjoint union of its left cosets; g_1 and g_2 are in the same left coset exactly when there is an element h in H with $g_1 \cdot h = g_2$. The set of left cosets is denoted by G/H. There is a natural map π from G onto G/H that takes an element in G to the coset containing it. A subgroup H is a *normal* subgroup if, for all g in G and h in H, $g \cdot h \cdot g^{-1}$ is in H. In this case G/H gets the structure of a group, in such a way that the natural map π: $G \to G/H$ is a homomorphism of groups.

The identity element in a group G is usually denoted by e, or e_G if there is chance of confusion. If φ: $G \to G'$ is a homomorphism of groups, the *kernel* $N = \{g \in G \colon \varphi(g) = e_{G'}\}$ is a normal subgroup of G, denoted $\mathrm{Ker}(\varphi)$. Then φ determines a one-to-one homomorphism $\overline{\varphi}$: $G/\mathrm{Ker}(\varphi) \to G'$ such that

$\varphi = \overline{\varphi} \circ \pi$. If φ is surjective, then $\overline{\varphi}$ is an isomorphism. More generally, if N is any normal subgroup of G, a homomorphism from G/N to a group G' determines a homomorphism from G to G' such that N is contained in its kernel. The *image* $\varphi(G)$ of any homomorphism $\varphi: G \to G'$ is a subgroup of G', denoted $\mathrm{Im}(\varphi)$, and φ determines an isomorphism of $G/\mathrm{Ker}(\varphi)$ with $\mathrm{Im}(\varphi)$. If $N \subset G$ and $N' \subset G'$ are normal subgroups, and $\varphi: G \to G'$ is a homomorphism such that $\varphi(N) \subset N'$, then φ determines a homomorphism $\overline{\varphi}: G/N \to G'/N'$.

The set of homomorphisms from G to G' is denoted by $\mathrm{Hom}(G, G')$. So if N is a normal subgroup of G,

$$\mathrm{Hom}(G/N, G') \leftrightarrow \{\varphi \in \mathrm{Hom}(G, G'): \varphi(N) = e_{G'}\}.$$

An important normal subgroup of a group G is the *commutator subgroup*, denoted $[G, G]$. This consists of all finite products

$$g_1 h_1 g_1^{-1} h_1^{-1} \cdot g_2 h_2 g_2^{-1} h_2^{-1} \cdot \, . \, . \, . \, \cdot g_n h_n g_n^{-1} h_n^{-1},$$

for elements $g_1, h_1, g_2, h_2, \ldots, g_n, h_n$ in G. The normality of this subgroup comes from the identity $g \cdot (ab) \cdot g^{-1} = (g \cdot a \cdot g^{-1}) \cdot (g \cdot b \cdot g^{-1})$. If A is an abelian group, any homomorphism of G to A sends all commutators to the identity, so

$$\mathrm{Hom}(G/[G, G], A) \leftrightarrow \mathrm{Hom}(G, A).$$

We usually use an *additive* notation for the product in abelian groups, writing $g + h$ instead of $g \cdot h$, with the identity element denoted 0. The group of integers under addition, which is the infinite cyclic group, is denoted $\mathbb{Z}$. The abelian group with just one element is often denoted 0. If A is an abelian group, and X is any set, the set of functions from X to A has a natural structure of abelian group, with $(f + g)(x) = f(x) + g(x)$. In particular, if G is any group, the set of homomorphisms $\mathrm{Hom}(G, A)$ has the structure of an abelian group.

Exercise C.4. If $\varphi: G \to G'$ is a homomorphism of groups, show that the mapping from $\mathrm{Hom}(G', A)$ to $\mathrm{Hom}(G, A)$ that takes ψ to $\psi \circ \varphi$ is a homomorphism of abelian groups.

If A and B are abelian groups, the *direct sum* $A \oplus B$ is the set of pairs (a, b), with a in A and b in B, with addition defined by $(a, b) + (a', b') = (a + a', b + b')$. More generally, given any collection A_α of abelian groups, the *direct sum* $\oplus A_\alpha$ consists of collections $\{a_\alpha\}$, with a_α in A_α, with the condition that a_α can be nonzero for only finitely many α. Addition is defined component by component, as for two factors. For example, $\mathbb{Z}^n$ is the direct sum of n copies of $\mathbb{Z}$. The same definition, but without the restriction that only finitely many are nonzero, defines the *direct product*, denoted ΠA_α.

Exercise C.5. If an abelian group C contains subgroups A and B such that

every element of C can be written as a sum of an element in A and an element in B, and $A \cap B = \{0\}$, show that $A \oplus B$ is isomorphic to C.

Exercise C.6. For any collection A_α of abelian groups, and any abelian group B, construct an isomorphism

$$\mathrm{Hom}(\oplus A_\alpha, B) \cong \Pi\, \mathrm{Hom}(A_\alpha, B).$$

An abelian group A is a *free abelian group*, with *basis* $\{e_\alpha\}$, if every element in the group has a unique expression in the form $\Sigma n_\alpha e_\alpha$, for some integers n_α, with only finitely many n_α nonzero. If the number of elements in a basis is a finite number n, we say A is a free abelian group of *rank n*. As we will see below, this number is independent of choice of basis.

Exercise C.7. (a) If A and B are free abelian groups, show that $A \oplus B$ is free abelian, and if the ranks are finite, $\mathrm{rank}(A \oplus B) = \mathrm{rank}(A) + \mathrm{rank}(B)$. (b) If F is free abelian, and A is abelian, and $\varphi: A \to F$ is a surjective homomorphism, show that A is isomorphic to the direct sum of F and $\mathrm{Ker}(\varphi)$.

More generally, a set $\{e_\alpha\}$ of elements in an abelian group A is called *linearly independent* if no linear combination of them is zero, i.e., there is no set of integers $\{n_\alpha\}$, with only finitely many nonzero, but not all zero, such that $\Sigma n_\alpha e_\alpha = 0$. The maximum number of elements in a linearly independent set in A is called the *rank* of A. We will prove at the end of this section that any two maximal linearly independent sets have the same number of elements, at least when this number is finite.

Unlike the case with vector spaces, however, a maximal set of linearly independent elements in an abelian group need not generate the group. For example, a finite abelian group has no independent elements. Even for groups with no elements of finite order, however, it is not true:

Exercise C.8. Show that the rank of the abelian group $\mathbb{Q}$ of rational numbers is 1.

For any set X, the *free abelian group* on X, denoted $F(X)$, can be defined as the set of finite formal linear combinations $\Sigma n_x x$, with n_x integers, the sum over a finite subset of X. The addition is defined coordinate-wise: $\Sigma n_x x + \Sigma m_x x = \Sigma(n_x + m_x)x$. More precisely, define $F(X)$ to be the set of functions from X to $\mathbb{Z}$ that are zero except on a finite subset of X. This is an abelian subgroup of the abelian group of functions from X to $\mathbb{Z}$. To the function $f: X \to \mathbb{Z}$ is associated the expression $\Sigma f(x)x$. The element x corresponds to the function that takes x to 1 and all other elements of X to zero. These elements form a basis for $F(X)$.

For any abelian group A, and any function φ from a set X to A, there is a unique homomorphism from $F(X)$ to A that takes $\Sigma n_x x$ to $\Sigma n_x \varphi(x)$. In particular, if $\varphi: X \to Y$ is any function, it determines a homomorphism from

$F(X)$ to $F(Y)$, taking $\Sigma n_x x$ to $\Sigma n_x \varphi(x)$, or taking $n_1 x_1 + \ldots + n_r x_r$ to $n_1\varphi(x_1) + \ldots + n_r\varphi(x_r)$.

Exercise C.9. If $\varphi: X \to Y$ is one-to-one, show that $F(X) \to F(Y)$ is one-to-one, and if $\varphi: X \to Y$ is surjective, show that $F(X) \to F(Y)$ is surjective.

If A is an abelian group, then the set of homomorphisms $\mathrm{Hom}(A, \mathbb{R})$ from A to $\mathbb{R}$ forms a real vector space, with scalar multiplication by the rule $(rf)(x) = r \cdot f(x)$ for f in $\mathrm{Hom}(A, \mathbb{R})$, r a real number, and x in A. If $\varphi: A \to A'$ is a homomorphism of abelian groups, then there is a linear mapping $\varphi^*: \mathrm{Hom}(A', \mathbb{R}) \to \mathrm{Hom}(A, \mathbb{R})$ of vector spaces, defined by the formula $\varphi^*(f) = f \circ \varphi$.

Lemma C.10. *If $\varphi: A \to A'$ is one-to-one, then φ^* is surjective.*

Proof. We need some preliminaries. There exists a set $\mathcal{B} = \{x_\alpha : \alpha \in \mathcal{A}\}$ of elements in A', such that:

(i) no finite linear combination $\Sigma n_\alpha x_\alpha$ with integer coefficients is in the image of φ unless all n_α are zero; and
(ii) $\mathcal{B}$ is maximal with this property.

This is a consequence of Zorn's lemma, exactly as in the proof that every vector space has a basis. Of course, there may be many such sets $\mathcal{B}$, but we fix one. It follows that for any element x in A', there is a nonzero integer n and integers n_α, all zero except for finitely many, so that $nx - \Sigma n_\alpha x_\alpha$ is in $\varphi(A)$; otherwise one could enlarge $\mathcal{B}$ by adding x to it. Therefore for any x in A' there is at least one equation of the form

$$nx = \Sigma n_\alpha x_\alpha + \varphi(y) \tag{C.11}$$

with y in A, n not 0.

Given f in $\mathrm{Hom}(A, \mathbb{R})$, we define g in $\mathrm{Hom}(A', \mathbb{R})$ by setting

$$g(x) = \frac{1}{n}(f(y)),$$

for any integer $n \neq 0$ and $y \in A$ so that (C.11) holds. To see that g is well defined, suppose also

$$mx = \Sigma m_\alpha x_\alpha + \varphi(z),$$

with $z \in A$ and $m \neq 0$. Then

$$\begin{aligned} \Sigma(mn_\alpha - nm_\alpha)x_\alpha &= (mnx - m\varphi(y)) - (nmx - n\varphi(z)) \\ &= \varphi(nz) - \varphi(my) = \varphi(nz - my). \end{aligned}$$

By (i), this element must be zero, and since φ is one-to-one, nz must be equal to my. Therefore,

$$\frac{1}{m}(f(z)) = \frac{1}{mn}(f(nz)) = \frac{1}{mn}(f(my)) = \frac{1}{mn}(mf(y)) = \frac{1}{n}(f(y)),$$

as required.

The proof that g is a homomorphism is similar, for if x and x' are two elements of A', write

$$nx = \Sigma n_\alpha x_\alpha + \varphi(y) \quad \text{and} \quad mx' = \Sigma m_\alpha x_\alpha + \varphi(z).$$

Then $mn(x \pm x') = \Sigma(mn_\alpha \pm nm_\alpha)x_\alpha + \varphi(my \pm nz)$, so

$$g(x \pm x') = \frac{1}{mn}(f(my \pm nz)) = \frac{1}{n}f(y) \pm \frac{1}{m}f(z) = g(x) \pm g(x').$$

And, by definition, if $x = \varphi(y)$, then $g(x) = f(y)$, so $\varphi^*(g) = f$. □

Corollary C.12. *Suppose S is a set with n elements in an abelian group A, and S is a maximal set of linearly independent elements. Then the dimension of Hom(A,* $\mathbb{R}$*) is equal to n. In particular, any two maximal sets of linearly independent elements in A have the same number of elements.*

Proof. Let F be the free abelian group on S, and $\varphi: F \to A$ the natural map. The linear independence of S assures that φ is one-to-one. The maximality of S assures that the set $\mathcal{B}$ considered in the proof of the lemma is empty. The proof of the lemma shows that the map φ^* from $\text{Hom}(A, \mathbb{R})$ to $\text{Hom}(F, \mathbb{R})$ is an isomorphism. The functions that take value 1 on a given element in S, and value 0 on the other elements, give a basis of $\text{Hom}(F, \mathbb{R})$ with n elements, and this shows that the dimension of $\text{Hom}(A, \mathbb{R})$ is n. Note that if S were a maximal set of linearly independent elements that were infinite, the same argument shows that $\text{Hom}(A, \mathbb{R}) \cong \text{Hom}(F, \mathbb{R})$ is infinite dimensional. □

Exercise C.13. Show conversely that if $\text{Hom}(A, \mathbb{R})$ has finite dimension n, then A has rank n.

Exercise C.14. If $B \to C$ is surjective, with kernel A, show that if two of the three abelian groups A, B, and C have finite ranks, so does the third, and $\text{rank}(B) = \text{rank}(A) + \text{rank}(C)$.

Given homomorphism $\varphi: A \to B$ and $\psi: B \to C$, one says that the sequence $A \xrightarrow{\varphi} B \xrightarrow{\psi} C$ is *exact*, or *exact at B* if the image of φ is equal to the kernel of ψ. To say that the sequence $0 \to A \to B$ is exact is the same as saying the map from A to B is one-to-one, and to say that $A \to B \to 0$ is exact is the same as saying the map from A to B is surjective.

Problem C.15. If $A \to B \to C$ is exact at B, show that the dual sequence $\mathrm{Hom}(C, \mathbb{R}) \to \mathrm{Hom}(B, \mathbb{R}) \to \mathrm{Hom}(A, \mathbb{R})$ is exact at $\mathrm{Hom}(B, \mathbb{R})$.

A sequence $A_0 \to A_1 \to \ldots \to A_n \to A_{n+1}$ of abelian groups and homomorphisms between them is called *exact* if it is exact at each of the groups A_i, for $1 \leq i \leq n$.

Problem C.16. Show that if $0 \to A_1 \to \ldots \to A_n \to 0$ is exact, and each of the abelian groups A_i has finite rank, then

$$\sum_{i=1}^{n} (-1)^i \operatorname{rank}(A_i) = 0 .$$

C3. Polynomials; Gauss's Lemma

For any field K the ring of polynomials $K[X]$ in a variable X over K is a *unique factorization domain*. In fact, every nonzero P in $K[X]$ has a unique factorization $P = a \cdot \Pi P_i^{n_i}$, with a in K and each P_i an irreducible polynomial that is *monic*.[9] This follows from the fact that one has a division algorithm for polynomials, just as one has for integers. In particular, any finite collection of nonzero polynomials has a greatest common divisor, which is unique if, in addition, it is required to be monic. The quotient field of $K[X]$, consisting of all ratios P/Q, $Q \neq 0$, is denoted $K(X)$.

The ring of polynomials $K[X, Y]$ in two variables X and Y is a subring of the ring $K(X)[Y]$, which, by what we have just seen, is a unique factorization domain.

Lemma C.17 (Gauss). *Let F be a polynomial in $K[X, Y]$. If F is irreducible in $K(X)[Y]$, then F is irreducible in $K[X, Y]$.*

Proof. Given F in $K[X, Y]$, write $F = a_0(X) + a_1(X)Y + \ldots + a_n(X)Y^n$, with $a_i(X)$ in $K[X]$. The greatest common divisor of $a_0(X), \ldots, a_n(X)$ is called the *content* of F, and denoted $c(F)$. Call F *primitive* if $c(F) = 1$.

We show first that the product of two primitive polynomials is also primitive. To see this, suppose $F = a_0 + a_1 Y + \ldots + a_n Y^n$ and $G = b_0 + b_1 Y + \ldots + b_m Y^m$ are primitive. Suppose a nonconstant polynomial $p = p(X)$ divides all the coefficients of $F \cdot G$. Take the minimal i and j such that p does not divide a_i and b_j. Then the coefficient of Y^{i+j} in $F \cdot G$ has the form

$$a_i b_j + a_{i+1} b_{j-1} + \ldots + a_{i+j} b_0 + a_{i-1} b_{j+1} + \ldots + a_0 b_{i+j} .$$

[9] A monic polynomial is one of the form $X^m + a_1 X^{m-1} + \ldots + a_m$.

Since all the terms but the first are divisible by p, and the first is not, this is a contradiction.

It follows from this that for any two polynomials F and G in $K[X, Y]$, $c(F \cdot G) = c(F) \cdot c(G)$. To see this, write $F = c(F) \cdot F_1$, $G = c(G) \cdot G_1$, with F_1 and G_1 primitive. Then $F \cdot G = c(F) \cdot c(G) \cdot F_1 \cdot G_1$, and $F_1 \cdot G_1$ is primitive, from which it follows that the content of $F \cdot G$ is $c(F) \cdot c(G)$.

Given any nonzero G in $K(X)[Y]$, one can write $G = g \cdot G_1$, with g in $K(X)$ and G_1 a primitive polynomial in $K[X, Y]$. Now suppose F is an irreducible polynomial in $K[X, Y]$, and that F factors in $K(X)[Y]$ into $G \cdot H$, with both G and H of positive degree in Y. Write $G = g \cdot G_1$, $H = h \cdot H_1$, with G_1 and H_1 primitive, and $g = p/q$, $h = r/s$, with $p, q, r, s \in K[X]$. Then $q \cdot s \cdot F = p \cdot r \cdot G_1 \cdot H_1$ in $K[X, Y]$. It follows that $q \cdot s \cdot c(F) = p \cdot r$. Hence $F = c(F) \cdot G_1 \cdot H_1$, which contradicts the irreducibility of F in $K[X, Y]$. □

Exercise C.18. Show that, for any field K, $K[X, Y]$ is a unique factorization domain. Generalize to polynomials in n variables.

If P is a polynomial in $K[X]$, the residue class ring $K[X]/(P)$ is the set of equivalence classes of polynomials in $K[X]$, two being equivalent when their difference is divisible by P. The residue classes $K[X]/(P)$ have the structure of a ring so that the natural map $K[X] \to K[X]/(P)$ is a homomorphism of rings.

Lemma C.19. *If P has degree n, then the images of $1, X, \ldots, X^{n-1}$ form a basis for $K[X]/(P)$ over K.*

Proof. These elements span, since, by dividing by P, any polynomial is equivalent to a polynomial of degree less than n. They are linearly independent, since no nonzero polynomial of degree less than n is divisible by P. □

Exercise C.20. Show that $K[X]/(P)$ is a field if and only if P is irreducible.

APPENDIX D

On Surfaces

D1. Vector Fields on Plane Domains

The object of this section is to prove Lemmas 7.10 and 7.11; we refer to Chapter 7 for notation. Suppose $\varphi: U \to U'$ is a diffeomorphism from one open set in the plane onto another. If $\varphi(x,y) = (u(x,y), v(x,y))$ in coordinates, at any point P in U, we have the Jacobian matrix

$$J_{\varphi,P} = \begin{bmatrix} \dfrac{\partial u}{\partial x}(P) & \dfrac{\partial u}{\partial y}(P) \\ \dfrac{\partial v}{\partial x}(P) & \dfrac{\partial v}{\partial y}(P) \end{bmatrix},$$

which we regard as a linear mapping from vectors in $\mathbb{R}^2$ to vectors in $\mathbb{R}^2$ (see Appendix C). If V is a continuous vector field in U, the vector field $\varphi_* V$ in U' is defined by the formula

$$(\varphi_* V)(P') = J_{\varphi,P}(V(P)),$$

where P is the point in U mapped to P' by φ, i.e., $P = \varphi^{-1}(P')$. If V has singularities in the set Z, $\varphi_* V$ will have singularities in $\varphi(Z)$.

Lemma D.1. $\text{Index}_{\varphi(P)}(\varphi_* V) = \text{Index}_P V$.

Proof. There is no loss in generality by assuming that P and P' are the origin 0, that U is a disk containing the origin, and that V is not zero in $U \setminus \{0\}$.

Let J be the Jacobian of φ at 0. Our first goal is to show that

$$\text{(D.2)} \qquad \text{Index}_0(\varphi_* V) = \text{Index}_0(J_* V).$$

This will reduce the problem to the easier case of a linear mapping. We want to construct a homotopy from φ to J. Define

$$K: U \times [0,1] \to \mathbb{R}^2, \qquad Q \times t \mapsto \begin{cases} \dfrac{1}{t}\varphi(t \cdot Q), & 0 < t \leq 1, \\ J(Q), & t = 0. \end{cases}$$

Claim D.3. *This mapping K is $\mathscr{C}^\infty$.*

To prove this claim, we use Lemma B.9, replacing U by a smaller disk if necessary, so we can write

$$\varphi(x,y) = (xu_1(x,y) + yu_2(x,y), xv_1(x,y) + yv_2(x,y)),$$

with $\mathscr{C}^\infty$ functions u_1, u_2, v_1, and v_2. Then

$$K((x,y) \times t) = (xu_1(tx,ty) + yu_2(tx,ty), xv_1(tx,ty) + yv_2(tx,ty))$$

for all $0 \leq t \leq 1$, and this expression is clearly $\mathscr{C}^\infty$.

Now $H(Q \times t) = (K_t)_*(V)$ gives a homotopy from J_*V to φ_*V in the sense of Exercise 7.3, and (D.2) follows from that exercise.

We are therefore reduced to showing that $\text{Index}_0(J_*V) = \text{Index}_0(V)$ for any invertible linear mapping J. Now we use Lemma C.2 to know that there is a path in the space of invertible matrices from J either to the identity matrix I, or to the matrix $I' = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$. If such a path is given by a formula $t \mapsto J_t$, $a \leq t \leq b$, then the homotopy $H(Q \times t) = (J_t)_*(V)$ gives a homotopy from J_*V to I_*V or to I'_*V, and the same exercise shows that the index doesn't change. Of course $I_*V = V$, so all that remains is to prove that $\text{Index}_0(I'_*V) = \text{Index}_0(V)$.

If $V(x,y) = (p(x,y), q(x,y))$, then by the definition of I'_*,

$$(I'_*V)(x,y) = (p(x,-y), -q(x,-y)).$$

So one is reduced to the elementary problem of showing that if $F(x,y) = (p(x,y), q(x,y))$, and $R(x,y) = (x, -y)$, the mappings $R \circ F \circ R$ and F, when restricted to a small circle, have the same winding number around the origin. This is easy to do directly from the definition of winding number, and we leave the details as an exercise. (This is also special case of the fact that the degree of a composite of mappings of circles is the product of the degrees of the mappings, see Problem 3.27. In this way one can argue directly with any linear mapping J, since $J_*V = J \circ V \circ J^{-1}$.) □

Now we consider the other lemma from Chapter 7.

Lemma D.4. *Suppose V and W are continuous vector fields with no singularities on an open set U containing a point P. Let $D \subset U$ be a closed disk centered at P. Then there is a vector field $\tilde{V}$ with no singularities on U such that (i) $\tilde{V}$ and V agree on $U \setminus D$; (ii) $\tilde{V}$ and W agree on some neighborhood of P.*

Proof. Suppose first that the dot product $V(P) \cdot W(P)$ is positive. Shrinking the disk, we may assume that $V(Q) \cdot W(Q) > 0$ for all Q in D. As in Step 2 of the proof of Proposition B.10 (see Exercise B.14), there is a $\mathscr{C}^\infty$ function ρ that is identically 1 in a neighborhood of P, and identically 0 outside D, and taking values in $[0, 1]$. Let

$$\tilde{V}(Q) = (1 - \rho(Q))V(Q) + \rho(Q)W(Q).$$

Then $V(Q) \cdot \tilde{V}(Q) > 0$ for all Q in U, so $\tilde{V}$ has no singularities, and conditions (i) and (ii) are clear.

For the general case, it therefore suffices to find a vector field V_1 with no singularities that agrees with V outside D, and such that $V_1(P) \cdot W(P)$ is positive. This can be done by rotating V inside D. With the same function ρ, and ϑ the angle from the vector $V(P)$ to the vector $W(P)$, we can take

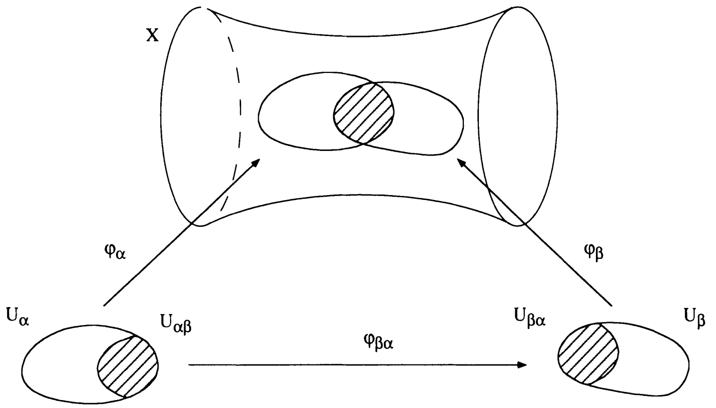
$$V_1(Q) = \begin{bmatrix} \cos(\rho(Q)\vartheta) & -\sin(\rho(Q)\vartheta) \\ \sin(\rho(Q)\vartheta) & \cos(\rho(Q)\vartheta) \end{bmatrix} \cdot V(Q). \qquad \square$$

D2. Charts and Vector Fields

We start with a brief definition of a (smooth) surface X, and define what we mean by a vector field on X and the index of a vector field at a point on X. A *surface X with an atlas of charts* is a Hausdorff topological space, equipped with a collection of homeomorphisms

$$\varphi_\alpha \colon U_\alpha \to \varphi_\alpha(U_\alpha) \subset X,$$

with U_α open in the plane $\mathbb{R}^2$, and $\varphi_\alpha(U_\alpha)$ open in X; the α are in some index set. The surface X should be the union of these open sets $\varphi_\alpha(U_\alpha)$. Let $U_{\alpha\beta} = \varphi_\alpha^{-1}(\varphi_\alpha(U_\alpha) \cap \varphi_\beta(U_\beta))$. These charts determine *change of coordinate mappings* $\varphi_{\beta\alpha} = \varphi_\beta^{-1} \circ \varphi_\alpha$, which are homeomorphisms from $U_{\alpha\beta}$ to $U_{\beta\alpha}$.



To give X a differentiable, or smooth, structure, the requirement is that these changes of coordinates $\varphi_{\beta\alpha}$ should be $\mathscr{C}^\infty$ mappings for all α and β. One then has a notion of a *differentiable function* on an open subset U of X: it is function f such that $f \circ \varphi_\alpha$ is differentiable on $\varphi_\alpha^{-1}(U) \cap U_\alpha$ for all α.

Another collection of charts $\{\psi_{\alpha'}: U_{\alpha'} \to X\}$ is said to be equivalent to this one if all the changes of coordinates from one to the other are all $\mathscr{C}^\infty$, i.e., all $\varphi_\alpha^{-1} \circ \psi_{\alpha'}$ are $\mathscr{C}^\infty$ where defined. We say that this collection defines the same surface. More precisely, a *smooth* (or $\mathscr{C}^\infty$) *surface* is the topological space X together with an equivalence class of families of charts. For the sphere S^2, the two mappings φ and ψ we obtained from stereographic projection in §7c form a family of charts. Stereographic projection from other points gives charts that are compatible in this sense.

If f is a $\mathscr{C}^\infty$ function on some open set in $\mathbb{R}^3$, and X is the locus where $f(x,y,z) = 0$ and $\text{grad}(f) \neq 0$, then X is a smooth surface. If for example $(\partial f/\partial z)(P) \neq 0$, the implicit function theorem says that projection from X to the xy-plane is locally one-to-one near P, and the inverse to this projection provides a chart near P.

If X is a surface given by charts as above, a *vector field* V on X can be defined as a compatible collection of vector fields V_α in each of the coordinate neighborhoods U_α. The compatibility means that

$$(\varphi_{\beta\alpha})_*(V_\alpha|_{U_{\alpha\beta}}) = V_\beta|_{U_{\beta\alpha}},$$

where $(\varphi_{\beta\alpha})_*$ is defined as in the preceding section. The vector field has a singularity at a point P in X if, with some α such that $\varphi_\alpha(P_\alpha) = P$, the corresponding V_α has a singularity at P_α in U_α. By Lemma D.1 the index of V_α at P_α is independent of choice of coordinate chart. This index is defined to be the *index* of V at P.

The surface X is *orientable* if it has an atlas of charts such that all the determinants of the Jacobians of the change of coordinate mappings are positive. An orientation is a choice of such an atlas, with two atlases defining

the same orientation if all changes of coordinates from one to the other have positive determinants of Jacobians. An orientable surface has two orientations, with two orientations defining the opposite orientation if all changes of coordinates from one to the other have negative Jacobian determinants.

D3. Differential Forms on a Surface

If $\varphi: U \to U'$ is a diffeomorphism from one open set in the plane to another, and $\omega = f\,dx + g\,dy$ is a $\mathscr{C}^\infty$ 1-form on U', one can define a *pull-back* 1-form $\varphi^*\omega$ on U by the formula

$$\begin{aligned}\varphi^*(f\,dx + g\,dy) &= f(\varphi(x,y)) \cdot \left(\frac{\partial u}{\partial x}(x,y)\,dx + \frac{\partial u}{\partial y}(x,y)\,dy\right)\\ &\quad + g(\varphi(x,y)) \cdot \left(\frac{\partial v}{\partial x}(x,y)\,dx + \frac{\partial v}{\partial y}(x,y)\,dy\right)\\ &= \left((f\circ\varphi)\cdot\frac{\partial u}{\partial x} + (g\circ\varphi)\cdot\frac{\partial v}{\partial x}\right)dx\\ &\quad + \left((f\circ\varphi)\cdot\frac{\partial u}{\partial y} + (g\circ\varphi)\cdot\frac{\partial v}{\partial y}\right)dy,\end{aligned}$$

where $\varphi(x,y) = (u(x,y), v(x,y))$. Similarly, if $\omega = h\,dx\,dy$ is a 2-form on U', the *pull-back* 2-form $\varphi^*\omega$ on U is defined by the formula

$$\varphi^*(h\,dx\,dy) = (h\circ\varphi)\cdot\left(\frac{\partial u}{\partial x}\frac{\partial v}{\partial y} - \frac{\partial u}{\partial y}\frac{\partial v}{\partial x}\right)dx\,dy.$$

Exercise D.5. (a) If $\psi: U' \to U''$ is a diffeomorphism, and ω is a 1-form or 2-form on U'', show that $\varphi^*(\psi^*\omega) = (\psi\circ\varphi)^*(\omega)$. (b) Show that $\varphi^*(df) = d(\varphi^*f)$ for a $\mathscr{C}^\infty$ function f on U', and $\varphi^*(d\omega) = d(\varphi^*\omega)$ for ω a 1-form on U'.

Given a surface X with an atlas of charts $\varphi_\alpha: U_\alpha \to X$ as in §D2, a function (or 0-form) is given by a collection of functions f_α on U_α such that they agree on the overlaps: $f_\alpha = f_\beta \circ \varphi_{\beta\alpha}$ on $U_{\alpha\beta}$. Define a *one-form* ω on X to be a collection of 1-forms ω_α on U_α that agree on the overlaps, i.e., such that

$$\omega_\alpha|_{U_{\alpha\beta}} = (\varphi_{\beta\alpha})^*(\omega_\beta|_{U_{\beta\alpha}})$$

for all pairs α and β. A *two-form* is defined likewise, taking the ω_α to be 2-forms on U_α.

If f is a $\mathscr{C}^\infty$ function (or *zero-form*) on X, its *differential* df is a 1-form, defined to be the 1-form $d(f\circ\varphi_\alpha)$ on U_α. Similarly, if ω is a 1-form on X, given by 1-forms ω_α on U_α, the *differential* of ω is the 2-form $d\omega$ defined to be the 2-form $d\omega_\alpha$ on U_α.

Exercise D.6. (a) Verify that these formulas define 1-forms and 2-forms on X. (b) Verify that d is linear, i.e., $d(r_1\omega_1 + r_2\omega_2) = r_1 d(\omega_1) + r_2 d(\omega_2)$ for real numbers r_1 and r_2 and 0-forms or 1-forms ω_1 and ω_2. (c) Verify that $d(df) = 0$. (d) Show that for $k = 0$, 1, and 2, a k-form for one atlas determines a k-form for any other atlas, and that this is compatible with the definition of differential.

There is also a *wedge product* $\wedge$ that takes two 1-forms ω and μ and produces a 2-form $\omega \wedge \mu$. For an open set U in the plane, if $\omega = f\,dx + g\,dy$, and $\mu = h\,dx + k\,dy$, for f, g, h, and k $\mathscr{C}^\infty$ functions on U, then

$$\omega \wedge \mu \;=\; (f\,dx + g\,dy) \wedge (h\,dx + k\,dy) = (f \cdot k - g \cdot h)\,dx\,dy\,.$$

If ω and μ are 1-forms on X given on U_α by ω_α and μ_α, respectively, then $\omega \wedge \mu$ is the 2-form given on U_α by $\omega_\alpha \wedge \mu_\alpha$, with $\omega_\alpha \wedge \mu_\alpha$ defined by the displayed formula. If f is a function and ω is a 1-form (or 2-form), then $f \cdot \omega$ (defined locally by $f_\alpha \cdot \omega_\alpha$) is a 1-form (or 2-form).

Exercise D.7. (a) Verify that $\omega \wedge \mu$ is a 2-form. (b) Verify the following properties of the wedge product:

(i) $$(f_1\omega_1 + f_2\omega_2) \wedge \mu \;=\; f_1(\omega_1 \wedge \mu) + f_2(\omega_2 \wedge \mu)$$

for ω_1, ω_2, and μ 1-forms, and f_1 and f_2 functions;

(ii) $$\mu \wedge \omega \;=\; -\omega \wedge \mu$$

for μ and ω 1-forms; and

(iii) $$d(f \cdot \mu) \;=\; (df) \wedge \mu + f \cdot d\mu$$

for f a function and μ a 1-form.

In fact, all the results of this appendix generalize from two to n dimensions, leading to the notion of a smooth manifold of dimension n, a vector field on a manifold, the index of a vector field, an orientation of a manifold, k-forms on a manifold ($0 \le k \le n$), with differential from k-forms to $(k+1)$-forms, wedge products of k-forms and l-forms being $(k+l)$-forms, with similar properties.

APPENDIX E

Proof of Borsuk's Theorem

This appendix contains a proof of Borsuk's theorem as stated in §23c. It assumes a knowledge of §23a and §23b.

So far in this book we have considered chains $\Sigma n_i \Gamma_i$ with integer coefficients n_i. In fact, one can use coefficients in any abelian group G, and one gets chains $C_k(X;G)$, cycles $Z_k(X;G)$, and boundaries $B_k(X;G)$, so homology groups $H_k(X;G) = Z_k(X;G)/B_k(X;G)$. All the formal properties proved about ordinary homology groups extend without change to these groups, and many of the calculations are similar. For example, the Mayer–Vietoris theorem is true without change.

It is often useful to look at coefficients in $\mathbb{Z}/p\mathbb{Z}$, the integers modulo a prime p. In this case there is a natural homomorphism from each $H_k(X)$ to $H_k(X;\mathbb{Z}/p\mathbb{Z},)$ obtained by reducing all coefficients modulo p. This gives homomorphisms

$$H_k(X)/pH_k(X) \to H_k(X;\mathbb{Z}/p\mathbb{Z}).$$

These homomorphisms are isomorphisms if X is a sphere, as one sees by tracing through the Mayer–Vietoris argument computing the homology of a sphere. The following exercise shows that it is not always an isomorphism, however.

Exercise E.1. (a) Show that the above map is an isomorphism if X is a compact oriented surface. (b) Show that, for X the real projective plane, or any compact surface, $H_2(X;\mathbb{Z}/2\mathbb{Z}) = \mathbb{Z}/2\mathbb{Z}$.

These homology groups are particularly useful when $p = 2$. In fact, every-

thing becomes a little easier, since one can ignore all signs entirely. Since these are the only ones we will consider here, we denote them by $\overline{H}_k X$:

$$\overline{H}_k X = H_k(X; \mathbb{Z}/2\mathbb{Z}).$$

Similarly, we denote by $\overline{C}_* X$ the chain complex $C_*(X; \mathbb{Z}/2\mathbb{Z})$.

It follows from the isomorphism $H_n(S^n)/2H_n(S^n) \cong \overline{H}_n(S^n)$ that for any continuous map $f\colon S^n \to S^n$, the degree of f is odd if the homomorphism $f_*\colon \overline{H}_n(S^n) \to \overline{H}_n(S^n)$ is not zero, and even if f_* is zero. We denote the antipodal map by

$$T\colon S^n \to S^n.$$

Let $\pi\colon S^n \to \mathbb{RP}^n$ be the two-sheeted covering map that identifies antipodal points. We will prove Borsuk's theorem by using the chain complex and homology of these spaces and maps with coefficients in $\mathbb{Z}/2\mathbb{Z}$. For this we need a general lemma.

Let $\pi\colon X \to Y$ be any two-sheeted covering map, and let $T\colon X \to X$ be the map which interchanges the two points in $\pi^{-1}(P)$ for each P in Y. The corresponding map $\pi_*\colon \overline{C}_* X \to \overline{C}_* Y$ of chain complexes is surjective. In fact, if $\Gamma\colon I^k \to Y$ is a k-cube, there are exactly two k-cubes Λ_1 and $\Lambda_2 = T \circ \Lambda_1$ in X with $\pi \circ \Lambda_i = \Gamma$. This is proved exactly as we proved the lifting of homotopies in §11b. The mapping $\Gamma \mapsto \Lambda_1 + \Lambda_2$ determines a homomorphism $t_*\colon \overline{C}_* Y \to \overline{C}_* X$ of chain complexes, called the *transfer* (see Problem 18.26). A chain is in the kernel of π_* exactly when each cube $T \circ \Lambda$ occurs with the same coefficient as Λ. We therefore have an exact sequence of chain complexes

$$0 \to \overline{C}_* Y \xrightarrow{t_*} \overline{C}_* X \xrightarrow{\pi_*} \overline{C}_* Y \to 0. \tag{E.2}$$

This gives a long exact sequence in homology:

$$\ldots \to \overline{H}_k Y \xrightarrow{t_*} \overline{H}_k X \xrightarrow{\pi_*} \overline{H}_k Y \xrightarrow{\partial} \overline{H}_{k-1} Y \xrightarrow{t_*} \overline{H}_{k-1} X \to \ldots \quad .$$

Lemma E.3. *(i) Let $f\colon X \to X$ be a continuous map such that $f \circ T = T \circ f$, and let $g\colon Y \to Y$ be the map determined by the condition $g \circ \pi = \pi \circ f$. Then the diagram*

$$\begin{array}{ccccccccc} 0 & \longrightarrow & \overline{C}_* Y & \xrightarrow{t_*} & \overline{C}_* X & \xrightarrow{\pi_*} & \overline{C}_* Y & \longrightarrow & 0 \\ & & \downarrow g_* & & \downarrow f_* & & \downarrow g_* & & \\ 0 & \longrightarrow & \overline{C}_* Y & \xrightarrow{t_*} & \overline{C}_* X & \xrightarrow{\pi_*} & \overline{C}_* Y & \longrightarrow & 0 \end{array}$$

commutes.

(ii) Let $f\colon X \to X$ be a continuous map such that $f \circ T = f$, and let $g\colon Y \to Y$ be the map determined by the condition $g \circ \pi = \pi \circ f$. Then the diagram

$$\begin{array}{ccccccccc} 0 & \longrightarrow & \bar{C}_*Y & \xrightarrow{t_*} & \bar{C}_*X & \xrightarrow{\pi_*} & \bar{C}_*Y & \longrightarrow & 0 \\ & & \downarrow{\scriptstyle 0} & & \downarrow{\scriptstyle f_*} & & \downarrow{\scriptstyle g_*} & & \\ 0 & \longrightarrow & \bar{C}_*Y & \xrightarrow{t_*} & \bar{C}_*X & \xrightarrow{\pi_*} & \bar{C}_*Y & \longrightarrow & 0 \end{array}$$

commutes.

Proof. Both of these are straightforward from the definitions. The right squares commute by functoriality: $g_* \circ \pi_* = (g \circ \pi)_* = (f \circ \pi)_* = f_* \circ \pi_*$. For the left squares, let Γ be a cube in Y, Λ_1 and Λ_2 its two liftings, so $f_*t_*[\Gamma] = [f \circ \Lambda_1] + [f \circ \Lambda_2]$. In case (i), $f \circ \Lambda_1$ and $f \circ \Lambda_2$ are the two liftings of $g \circ \Gamma$, so $t_*g_*[\Gamma] = [f \circ \Lambda_1] + [f \circ \Lambda_2]$, as required. In case (ii), $f \circ \Lambda_2 = f \circ T \circ \Lambda_1 = f \circ \Lambda_1$, so $f_*t_*[\Gamma] = 2[f \circ \Lambda_1] = 0$. □

It follows from this lemma that the corresponding maps between the long exact homology sequences commute. We apply this now to $X = S^n$, $Y = \mathbb{RP}^n$, $\pi: X \to Y$ the covering, and T the antipodal map. The long exact sequence arising from (E.2) takes the form

$$0 \longrightarrow \bar{H}_nY \xrightarrow{t_*} \bar{H}_nX \xrightarrow{\pi_*} \bar{H}_nY \xrightarrow{\partial} \bar{H}_{n-1}Y \longrightarrow 0$$

$$0 \longrightarrow \bar{H}_{n-1}Y \xrightarrow{\partial} \bar{H}_{n-2}Y \longrightarrow 0 \longrightarrow \dots$$

$$\dots \longrightarrow 0 \longrightarrow \bar{H}_1Y \xrightarrow{\partial} \bar{H}_0Y \xrightarrow{t_*} \bar{H}_0X \xrightarrow{\pi_*} \bar{H}_0Y \longrightarrow 0.$$

Since $\bar{H}_nX = \mathbb{Z}/2\mathbb{Z}$, we must have $\bar{H}_nY \neq 0$, so $t_*: \bar{H}_nY \to \bar{H}_nX$ is an isomorphism. Hence $\pi_*: \bar{H}_nX \to \bar{H}_nY$ is zero, so $\partial: \bar{H}_nY \to \bar{H}_{n-1}Y$ is an isomorphism. Continuing, we see that $\partial: \bar{H}_iY \to \bar{H}_{i-1}Y$ is an isomorphism for all $i = 1, \dots, n$. In particular, $\bar{H}_iY = \mathbb{Z}/2\mathbb{Z}$ for $i = 0, \dots, n$.

Now suppose $f: S^n \to S^n$ is a map with $f \circ T = T \circ f$. Borsuk's theorem (Theorem 23.24(a)) states that the degree of f is odd, which is equivalent to saying that $f_*: \bar{H}_nX \to \bar{H}_nX$ is an isomorphism. Applying Lemma E.3(i), we have commuting diagrams

$$\begin{array}{ccc} \bar{H}_nY & \xrightarrow[\cong]{t_*} & \bar{H}_nX \\ \downarrow{\scriptstyle g_*} & & \downarrow{\scriptstyle f_*} \\ \bar{H}_nY & \xrightarrow[\cong]{t_*} & \bar{H}_nX, \end{array} \qquad \begin{array}{ccc} \bar{H}_iY & \xrightarrow[\cong]{\partial} & \bar{H}_{i-1}Y \\ \downarrow{\scriptstyle g_*} & & \downarrow{\scriptstyle g_*} \\ \bar{H}_iY & \xrightarrow[\cong]{\partial} & \bar{H}_{i-1}Y, \end{array}$$

the right diagrams valid for $i = 1, \dots, n$. Since $g_*: \bar{H}_0Y \to \bar{H}_0Y$ is an isomorphism, it follows from the right squares and induction on i that the homomorphism $g_*: \bar{H}_iY \to \bar{H}_iY$ is an isomorphism for every $i = 0, \dots, n$. Then the left square implies that $f_*: \bar{H}_nX \to \bar{H}_nX$ is an isomorphism as well, and this completes the proof.

Similarly, if $f: S^n \to S^n$ is a map with $f \circ T = f$, we have by (ii) of the lemma a commutative diagram

$$\begin{array}{ccc} \bar{H}_n Y & \xrightarrow[\cong]{t_*} & \bar{H}_n X \\ {\scriptstyle 0}\downarrow & & \downarrow{\scriptstyle f_*} \\ \bar{H}_n Y & \xrightarrow[\cong]{t_*} & \bar{H}_n X . \end{array}$$

This implies that $f_*: \bar{H}_n X \to \bar{H}_n X$ is zero, which means that f has even degree, and completes the proof of Theorem 23.24. □

The constructions of this section are part of a general development of P.A. Smith to study spaces equipped with periodic transformations like T. For a proof of Borsuk's theorem using simplicial approximations, see Armstrong (1983). For a proof using differential topology, see Guillemin and Pollack (1974).

Hints and Answers

Hints and/or answers are given for some of the exercises and problems, especially those used in the text, or those that are hard.

0.1. *Hint*: The answer depends only on the numbers of edges that emanate from each vertex. What happens to these numbers when you travel, erasing the edges as you travel over them? When do you get stuck at a vertex?

Answer: There is always an even number of vertices such that the number of edges emanating from the vertex is odd (if an edge has both ends at a vertex, it counts twice). If this number is greater than 2, the graph cannot be traced. If the number is 2, it can be traced, but only by starting at one of these, and (necessarily) ending at the other. If the number is 0, you can start anywhere, and will end at where you start. To see that one can do it under these conditions, one way is to make any trip, starting at an odd vertex if there are two such, continuing until you get stuck. Then make another trip, but adding a side trip along untraveled roads, until you (necessarily) get back to the old route at the same point. Each trip becomes longer, until the whole is traced.

0.2. See Chapter 8.

1.6. All but (vi).

1.7. Yes. Find such a function by integrating.

1.9. For the challenge, if P is in the closure of the points one can connect to P_0 by such an arc, take a disk D around P in U, take an are from P_0 to

a point inside D, and look at the first time the arc hits the boundary of D with inward pointing tangent; splice on to this arc (see §B2 for similar constructions) to get to any point inside D.

1.13. See Chapter 9 for more general results.

1.20. For the challenge, use the law of cosines.

2.3. Show that the derivative of $\vartheta(t)$ must be

$$\frac{-y(t)x'(t) + x(t)y'(t)}{x(t)^2 + y(t)^2}.$$

Or show that two such functions differ by a multiple of 2π, which must be 0 by continuity.

2.9. Consider neighborhoods given in polar coordinates by $\vartheta_1 < \vartheta < \vartheta_2$ and $r_1 < r < r_2$, with $\vartheta_2 - \vartheta_1 < 2\pi$.

2.13. See §B2 for the construction of such functions.

2.19. For formulas see Chapter 12, and use Problem 2.13.

2.22. Either argue directly, as in Appendix B, or use polar coordinates to map a rectangle onto the disk, and integrate the pull-back of the 1-form as in the first proof of Proposition 2.16.

2.24. Apply Green's formula (i) with $g = f$.

2.25. *Hint*: Apply Green's formula (ii), where R is the region inside the disk and outside a small disk around the point, with g of the form $a + b\log(r)$, where r is the distance from the center of the disk. Pass to the limit as the radius of the small disk approaches 0.

3.4. Use the definition. Choices of subdivision and sector U_i and ϑ_i for γ and P determine the same subdivision for $\gamma + v$, choices $U_i + v$ for sectors, and translated angle functions, so that the changes in angle along each piece are the same for each.

3.5. Use Exercise 2.9. See §11b for a generalization.

3.7. See Chapter 12 for formulas.

3.10. Apply the Lebesgue lemma to $\gamma \circ \varphi$, to obtain a subdivision $a' \le t_0' \le \ldots \le t_n' = b'$, such that $\gamma \circ \varphi$ maps each subinterval into a sector.

3.13. In the starshaped case, show that any closed path is homotopic to a constant path, and use Exercise 3.7.

3.15. Use Problem 3.14 to construct a homotopy between the lifted paths.

3.22. Use Problem 3.21 and Problem 3.14.

3.23. Use a homotopy $H(P \times s) = (1 - s)F(P) \pm sP$, which is a homotopy from F to the mapping $P \mapsto P$ or to $P \mapsto -P$.

3.25. Part (d) uses Problem 3.22.

3.27. The degree of $G \circ F$ is the sum of the degrees of F and G. For the proof, use Problem 3.26.

3.29. Use the fact that the group $\mathrm{GL}_2(\mathbb{R})$ of two by two invertible matrices has two connected components, cf. Appendix C.

3.30. Deform from the map to its linear approximation. See §D1.

3.31. See §19a for the local structure of general analytic mappings.

3.32. Compute the change in angle of F along the arcs between points in $F^{-1}(P')$.

4.6. If r were a retraction, $F(P) = -r(P)$ would have no fixed point.

4.7. Given a mapping f of Y to itself, consider the composite $i \circ f \circ r$, where i is the inclusion of Y in X and r is the retract.

4.8. (i). Note that (ii) and (iv) are homeomorphic.

4.11. Compare the restriction of f to S^1 with the identity mapping and the antipodal mapping, using the dog-on-a-leash theorem. Or look at fixed points of $x \mapsto \pm f(x)/|f(x)|$.

4.12. The unit vectors in the octant form a space homeomorphic to a disk, and, if no such vector is mapped to zero, then $F(P)/\|F(P)\|$ must have a fixed point.

4.13. Use the preceding exercise.

4.14. Look at the mapping $P \mapsto f(P) - P$, and use Exercise 4.11.

4.15. Show that f is homotopic to the antipodal map.

4.17. See Problem 3.23.

4.18. Map D^∞ to S^∞ by a formula $(a_0, a_1, \ldots) \mapsto (t, a_0, a_1, \ldots)$.

4.24. See the proof of Lemma 4.20.

4.27. See Lemma 4.21.

4.30. If not, do a spherical projection from a point not in the image.

4.31. If $f(P) \neq P$ for all P, f is homotopic to the antipodal map, while if $f(P) \neq P^*$ for all P, f is homotopic to the identity map.

4.38. Choose three arcs covering the circle without antipodal points, and look at their inverse images in the sphere.

4.39. Look at $A \cup B^*$, $B \cup C^*$, and $C \cup A^*$.

4.40. Tennis anyone?

5.4. If A is unbounded, the same proof shows that ω_P is exact. If A is bounded, the integral of ω_P around a large circle is nonzero.

5.13. Take $U = \mathbb{R}^2 \setminus A$, $V = \mathbb{R}^2 \setminus B$, and show that the image of δ and the kernel of δ each have dimension at least 1.

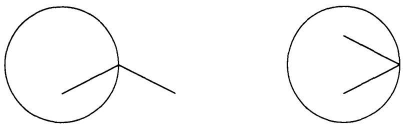
5.14. Divide the rectangle in half, with the intersection an interval. Argue as in Theorem 5.1, and use the fact (*) to know about the first cohomology groups of the complements.

5.16. Write $X = A \cup B$, with A and B homeomorphic to circles, and $A \cap B$ a point or homeomorphic to an interval.

5.21. Induct on e. Let A be the union of the vertices and $e - 1$ of the edges, and let B be the other (closed) edge, and set $U = \mathbb{R}^2 \setminus A$ and $V = \mathbb{R}^2 \setminus B$, arguing separately the cases when B has one endpoint or two, and, when two, whether they are in the same component of A or not.

5.22. Analyze the connected components of the complement as the edges are added.

5.23. Take $V = \mathbb{R}^2 \setminus X$, so $U \cup V = \mathbb{R}^2$ and $U \cap V = U \setminus X$.

5.25. 

5.26. Look at the image of a circle around the band, and the image of its complement.

5.27. For example, the situation should look locally—via a homeomorphism—like the two axes crossing at the origin.

5.28. Both follow from Corollary 5.18.

6.5. If $\gamma = \Sigma n_i \gamma_i$, then $\gamma = \Sigma n_i(\partial \Gamma_i)$, where $\Gamma_i(t, s) = (1 - s) \cdot \gamma_i(t) + s \cdot P_0$, where P_0 is the point with respect to which U is starshaped.

6.12. Use Theorem 6.11. See §9a for generalizations.

6.14. If Γ: $[0, 1] \times [0, 1] \to U$, the boundary of $F \circ \Gamma$ is $F_*(\partial \Gamma)$.

6.17. See Proposition 7.5 and Problem 7.8.

6.18. See Problem 3.23.

6.24. (b) A point on a circle is a retract but not a deformation retract.

6.25. Use Exercise 6.20 and Proposition 6.23 to show that r_* is the inverse isomorphism to i_*.

6.27. If P' is a point not in X', use Tietze to extend F to a continuous mapping from an open neighborhood U of X to $\mathbb{R}^2 \setminus \{P'\}$. Apply Theorem 6.11.

6.28. The converse is false! For a counterexample, see Problem 13.28.

7.2. Consider $(\mathfrak{R}(x+iy)^n, \mathfrak{T}(x+iy)^n)$ and $(\mathfrak{R}(x+iy)^n, -\mathfrak{T}(x+iy)^n)$, where $\mathfrak{R}$ and $\mathfrak{T}$ denote the real and imaginary parts.

7.3. If γ is a path around P as usual, $H \circ \gamma$ gives a homotopy in $\mathbb{R}^2 \setminus \{0\}$ from $V_0 \circ \gamma$ to $V_1 \circ \gamma$.

7.4. In the first case, consider the homotopy

$$H(t, s) = s \cdot V(\gamma_r(t)) + (1 - s) \cdot \rho(\gamma_r(t)) \cdot V(\gamma_r(t)),$$

$0 \le t \le 1$, $0 \le s \le 1$, where γ_r is a path around a small circle around P. For the second, compare V and $-V$.

7.8. If an infinite number of points P_i in Z have $W(\gamma, P_i) \neq 0$, such points lie in a bounded set, so they must have a limit point P. Since P cannot be in U, $W(\gamma, P) = 0$, and this contradicts Proposition 6.8.

7.9. Recall that f has a local maximum (resp. minimum) at P if the Hessian is positive and $(\partial^2 f/\partial x^2)(P) < 0$ (resp. $(\partial^2 f/\partial x^2)(P) > 0$); and f has a saddle point at P if the Hessian is negative. Use Problem 3.30.

7.13. Show that

$$(\varphi^* V)(x, y) = \begin{bmatrix} y^2 - x^2 & 2xy \\ -2xy & y^2 - x^2 \end{bmatrix} \cdot \begin{bmatrix} a + p(x, y) \\ b + q(x, y) \end{bmatrix},$$

with (a, b) a nonzero vector, and where $p(x, y)$ and $q(x, y)$ approach zero as $|x|$ and $|y|$ approach infinity. Restrict to a large circle, and do a homotopy (see Appendix D), to deform $\begin{bmatrix} a + p(x, y) \\ b + q(x, y) \end{bmatrix}$ first to $\begin{bmatrix} a \\ b \end{bmatrix}$, and then to $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$.

7.16. See the first picture in this section.

7.17. Look at $V(P) = f(P) - (f(P) \cdot P)P$.

8.4. Thinking of a horizontal doughnut with g holes, put a source on top between each of the holes, and a sink directly under each source.

8.5. #peaks + #valleys − #passes $= 2 - 2g$.

8.6. No.

8.10. Triangulate each of the polygons by putting a new vertex in its center.

8.12. Lift the vector field or triangulation to S^2.

8.13. 0.

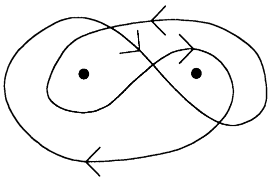
8.14. (b) A Klein bottle.

8.15. −1.

9.6. There is no homomorphism from $\mathbb{Z}^{n'}$ onto $\mathbb{Z}^n$ if $n > n'$, see Exercise C.14.

9.7. For (a), take a small circle γ_i around i for each $i \in \mathbb{N}$, and show that any closed 1-chain is homologous to a finite sum $\Sigma n_i \gamma_i$. The answer is the same for (b), since the spaces are homeomorphic, for example by the map $z \mapsto 1/z$ from $\mathbb{C} \setminus \{0\}$ to $\mathbb{C} \setminus \{0\}$.

9.9. The fact that the map is one-to-one follows from Theorem 6.11. For surjectivity, one can produce 1-cycles with arbitrary winding numbers around each K_i by Lemma 9.1, so the essential point is to show that the map from $H_1(U \setminus K)$ to $H_1 U$ is surjective. Take a grid so that no rectangle meets a point of K and a point not in U. If γ is a 1-cycle on U, we know γ is homologous to a sum of the form $\Sigma n_i \partial R_i$. Let $\gamma' = \Sigma_{R_i \cap K = \varnothing} n_i \partial R_i$. Then γ' is a 1-cycle on $U \setminus K$ that is homologous to γ on U. (See also Exercise 10.14.)

9.14. Consider the path  . See Exercise 11.14.

9.15. Use Corollary 9.12. If each γ_i is $\mathscr{C}^\infty$, so are all the constructions made in the proof that γ is a boundary.

9.16. Show that $\mathbb{R}^2 \setminus U$ is connected. Note that if V is any connected component of $\mathbb{R}^2 \setminus X$, then $\overline{V} \subset V \cup X$ and $\overline{V}$ meets X, so the union of any such V with X is connected.

9.17. For (a), take a subdivision and rectangles U_i as in the definition, but with the additional properties that each side of each U_i is of length at most 1, and the closure of U_i is contained in U. Let O_i be the point in the center of U_i. Given a point t_0, and an $\varepsilon > 0$, show that there is a $\delta > 0$ so that

$$|p(x,y,t) - p(x,y,t_0)| < \varepsilon/2n \quad \text{and} \quad |q(x,y,t) - q(x,y,t_0)| < \varepsilon/2n$$

for $|t - t_0| < \delta$. For each such t, let $f_{i,t}$ be the function on U_i so that $d(f_{i,t}) = \omega_t$ on U_i and such that $f_{i,t}(O_i) = 0$. Use the construction of Proposition 1.12, with the fact that the integrals are taken over segments of length at most $1/2$ (see Exercise B.8) to show that $|f_{i,t}(P) - f_{i,t_0}(P)| < \varepsilon/2n$ for all P in U_i, and for all $|t - t_0| < \delta$. Deduce that for $|t - t_0| < \delta$,

$$\begin{aligned} \left|\int_\gamma \omega_t - \int_\gamma \omega_{t_0}\right| &= \sum_{i=1}^{n} ((f_{i,t}(P_i) - f_{i,t_0}(P_i)) - (f_{i,t}(P_{i-1}) - f_{i,t_0}(P_{i-1}))) \\ &\le 2n \cdot (\varepsilon/2n) = \varepsilon . \end{aligned}$$

For (b), note that integral-valued locally constant functions are constant, and

they vanish if they are small.

9.20. 53π.

9.21. Let P_i be any point in A_i, and set $\omega = \Sigma\, \mathfrak{p}_i \omega_{P_i}$, where ω_{P_i} is the 1-form $(1/2\pi)\omega_{P_i,\vartheta}$ that measures change in angle around P_i.

9.22. For (b), use the equation displayed after Corollary 9.19, with $(m_1, m_2) = (1, 0)$ and $(0, 1)$.

9.23. Approximate $u(a + \Delta x + i\Delta y) - u(a)$ by $(\partial u/\partial x)(a)\Delta x + (\partial u/\partial y)(a)\,\Delta y$, and similarly for v.

9.25. For (b), take γ_ε a circle of radius ε around the singularity, and let ε approach 0.

9.28. See Problem 7.8.

9.31. See Exercise 7.4.

9.35. Use the dog-on-a-leash theorem (Theorem 3.11).

10.5. On a path-connected space, 0-cycles of degree zero are boundaries.

10.14. Apply Mayer–Vietoris, with $V = \mathbb{R}^2 \setminus K$. Use Corollary 9.4 to calculate $H_1 V$.

10.18. Identify S^2 with $\mathbb{R}^2 \cup \{\infty\}$. Suppose $X \setminus X \cap U = \{P, Q\}$, with P and Q in $\mathbb{R}^2$. Let $V = S^2 \setminus X$. Use Mayer–Vietoris to see that $U \cap V$ is disconnected exactly when $H_1 U \to H_1(S^2 \setminus \{P, Q\})$ is zero, or equivalently, when $W(\gamma, P) = W(\gamma, Q)$ for all 1-cycles γ on U.

10.19. See Problem C.16.

10.20. Use Corollary 9.4 to compute $H_1 U$, $H_1 V$, and $H_1(U \cup V)$; and compute the kernel of ∂. For the last part, argue as in the proof of Theorem 5.11, using the inclusion of $H_0(U \cap V)$ in $H_0(U) \oplus H_0(V)$ to show that if points P_0 and P_1 are in different connected components of U and V, then they are in different connected components of $U \cap V$.

10.21. Use Exercise 10.14 with MV(iii) and MV(iv).

10.22. Apply Alexander's lemma, with the compact sets X and $\partial D \cup B$.

10.23. If not, find points P_n and Q_n within distance $1/n$ that cannot be so joined, and apply the preceding problem to a limiting point P.

10.24. Let $\varepsilon_n = 1/n$, and take corresponding $\delta_n > 0$ from the preceding problem. Take a sequence of points Q_n in the same component as Q, with the distance from Q_n to P at most $\delta_n/2$. Connect Q_n to Q_{n+1} by a path in a disk of radius ε_n. Join these paths together, with the nth path defined on a subinterval of length $1/2^n$.

10.25. Use Problem 9.9 with MV(iii) and MV(iv).

10.28. Take $V = \mathbb{R}^2 \setminus K$, and use the isomorphism described after Exercise 9.21.

10.30. If not, use grids to find a closed path γ in $\mathbb{R}^2 \setminus \partial U$ that has different winding numbers about two points of ∂U. Note that $\partial U \subset K$. Since γ is connected and does not meet $\partial U = \overline{U} \setminus U$, γ must be contained in U or in $\mathbb{R}^2 \setminus \overline{U}$. If γ is contained in U, then γ has the same winding number about all points in K, since K is connected and contained in $\mathbb{R}^2 \setminus U$; this contradicts the fact that $\partial U \subset K$. If γ does not meet $\overline{U}$, since $\overline{U}$ is connected, the winding number of γ is constant around points in $\overline{U}$, contracting the fact that $\partial U \subset \overline{U}$.

If K is closed and connected, the same is true. To see it, let $\overline{K}$ be the closure of K in $S^2 = \mathbb{R}^2 \cup \{\infty\}$, let P be a point in U, and apply the preceding case to $\overline{K} \subset S^2 \setminus \{P\} \cong R^2$.

11.1. For (ii) you can use the identity $\exp(x + iy) = e^x(\cos(y) + i \sin(y))$.

11.4. Show that the set in X where the cardinality is n is open and closed.

11.10. Write an open rectangle as an increasing union of closed rectangles.

11.11. For (a), when X is locally connected, the evenly covered neighborhoods N can be taken to be connected, and if $p^{-1}(N)$ is a disjoint union of open sets N_α, each mapping homeomorphically to N, then these N_α are the connected components of $p^{-1}(N)$. (b) follows, since Y' will be a union of those N_α that it meets.

11.13. For the triangulation, use the lifting propositions to lift any triangulation of S to a triangulation of S'.

11.15. For (b), for $n \in \mathbb{Z}$ and (r, ϑ) in the right half plane, the action is given by $n \cdot (r, \vartheta) = (r, \vartheta + 2\pi n)$.

11.22. The map from $X \times G$ to Y by $x \times g \mapsto g \cdot s(x)$ is a G-isomorphism.

11.23. Use the preceding exercise.

11.24. It's enough to look where the covering is trivial, as in the proof of Lemma 11.5.

11.28. Given y in Y, take disjoint neighborhoods U_g of $g \cdot y$, one for each g in G, and let V be the intersection of the open sets $g^{-1} \cdot U_g$.

11.39. Given such an automorphism φ of S, define an automorphism of the covering by the formula $y * \gamma \mapsto \varphi(y) * \gamma$ for any $y \in S$ and any path γ starting at x. Show that this is independent of choices.

12.1. Use the same formulas as in the preceding displays, but for the second variable s.

12.3. For (a),

$$H(t,s) = \begin{cases} \sigma\left(\dfrac{4t}{1+s}\right), & 0 \le t \le 1/4(1+s), \\ \tau(4t-s-1), & 1/4(1+s) \le t \le 1/4(2+s), \\ \mu\left(\dfrac{4t-s-2}{2-s}\right), & 1/4(2+s) \le t \le 1. \end{cases}$$

12.5. To compute $\pi_1(S^1,(1,0))$, see Problem 3.14, which is an easy consequence of the propositions in Chapter 11.

12.6. Use the homotopy $H(v,s) = sx + (1-s)v$.

12.8. Cut the path into a finite number pieces that map into hemispheres, say, and replace each piece by a homotopic arc with the same endpoints, for example an arc along a great circle.

12.10. If σ and τ are loops at e, consider the homotopy $H(t,s) = \sigma(t) \cdot \tau(s)$.

12.12. Map (t,s) to

$$\begin{cases} (0, 1-2t), & 0 \le t \le 1/2(1-s), \\ \left(\dfrac{4t+2s-2}{3s+1}, s\right), & 1/2(1-s) \le t \le 1/4(s+3), \\ (1, 4t-3), & 1/4(s+3) \le t \le 1. \end{cases}$$

Follow this by h to achieve the homotopy.

12.15. (i), (ii), (iv), and (v) are equivalent; (iii), (vi), and (viii) are equivalent; (vii) and (ix) are equivalent.

12.18. For S^1, let $H((x_1,x_2) \times s) = $ rotation by $s \cdot \pi$ acting on (x_1,x_2). For larger n, use the same formulas for each successive pair of the $n+1$ coordinates on $S^n \subset \mathbb{R}^{n+1}$.

12.20. Show that it has a circle SO(2) as a deformation retract.

13.10. Take $X \subset \mathbb{R}^2$ to be the union of the lines $L(n) = \{1/n\} \times [0,1]$, for all positive integers n, and three lines $L = [-1,1] \times \{1\}$, $M = \{-1\} \times [0,1]$, and $N = [-1,0] \times \{0\}$. Let x be the point $(0,0)$. Take two copies X_1 and X_2 of X, and denote by subscripts the corresponding lines and points in X_1 and X_2. Take Y to be the disjoint union of X_1 and X_2, topologized as usual except near the points x_1 and x_2. For a disk U of radius $\varepsilon < 1$ about $(0,0)$ in the plane, define a neighborhood $U(x_i)$ in X_i by

$$U(x_1) = (N_1 \cap U) \cup \bigcup_{n \text{ odd}} (L(n)_1 \cap U) \cup \bigcup_{n \text{ even}} (L(n)_2 \cap U),$$

$$U(x_2) = (N_2 \cap U) \cup \bigcup_{n \text{ odd}} (L(n)_2 \cap U) \cup \bigcup_{n \text{ even}} (L(n)_1 \cap U).$$

This space Y is connected, and the natural map from Y to X is a covering map.

13.12. The proof is exactly the same as for the theorem.

13.17. By Corollary 13.16, the answer to (a) is $\mathbb{Z}/n\mathbb{Z}$. For (b), the fundamental group is the group of translations of the plane described in Exercise 11.25. The subgroup in (c) is generated by a and b^2.

13.21. For (1), note that $\gamma \cdot \alpha$ is homotopic to $(\gamma \cdot \beta) \cdot (\beta^{-1} \cdot \alpha)$. For (3), if $\gamma \cdot \alpha$ is homotopic to $\gamma' \cdot \beta$, with α and β paths in N from z to w, then $\alpha \cdot \beta^{-1}$ is homotopic to ε_z in X, so γ is homotopic to $\gamma \cdot (\alpha \cdot \beta^{-1})$, so to $(\gamma \cdot \alpha) \cdot \beta^{-1}$, so to $(\gamma' \cdot \beta) \cdot \beta^{-1}$, so to $\gamma' \cdot (\beta \cdot \beta^{-1})$, and so to γ'.

13.22. Take Y to be the union of $\mathbb{R}$ with a copy X_n of a clamshell—but without its outer circle—attached at each point n in $\mathbb{Z} \subset \mathbb{R}$, and map Y to X by wrapping $\mathbb{R}$ around the outer circle of X once between each integer. Take $Z \to Y$ to be a covering that is nontrivial over a different circle of each X_n.

13.27. For $\Leftarrow$, write γ as a boundary on U, and apply r to both sides.

13.28. Let γ be a path which, on $[0, 1/2]$ first goes around C_1 counterclockwise, then C_2 counterclockwise, then C_1 clockwise, then C_2 clockwise. On $[1/2, 3/4]$ it does the same but using C_3 and C_4, and so on, on intervals of length $1/2^n$ using the circles C_{2n-1} and C_{2n}. For all k there are homomorphisms from the fundamental group to the free group F_k with k generators, obtained by using the first k circles (see Problem 13.25 or §14d). The image of $[\gamma]$ in F_{2n} is a commutator of n elements, but not of fewer than n elements. So $[\gamma]$ cannot be the commutator of any finite number of elements.

13.29. $(r, \vartheta) \mapsto \log(r) + i\vartheta$.

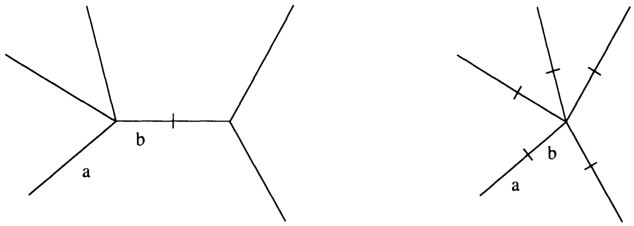
14.5. If Γ and Γ' were two such groups, the universal property for each would give maps from Γ to Γ' and from Γ' to Γ, and the two composites $\Gamma \to \Gamma' \to \Gamma$ and $\Gamma' \to \Gamma \to \Gamma'$ would be the identity maps by the uniqueness of such homomorphisms.

14.6. The subgroup of $\pi_1(X, x)$ that is generated by these images has the same universal property. It can be proved directly by subdividing the paths.

14.10. Take $G = \pi_1(U, x)$ and then $G = \pi_1(V, x)$.

14.13. See Problem 13.25 for the uniqueness.

14.15. If X' is X with an edge collapsed, map X' back to X as indicated:



The map is the identity except on edges adjacent to the vertex the edge is collapsed to. Send this vertex to the midpoint of the collapsing edge, and send half of each edge adjacent to this vertex to half of the edge that was collapsed, and spread the other half over the full edge. Check that the two composites are homotopic to the identity maps on X and X'.

14.17. For one point in a torus, the complement has a figure 8 as a deformation retract.

14.18. It is a free group on a countably infinite number of generators. Use Theorem 14.11.

14.19. The complement of an infinite discrete set in the plane is a covering space of the complement of a point in a torus.

14.20. The sphere with two handles can be obtained by joining the complements of disks in two tori. For another approach, see Chapter 17.

15.17. See Problem 9.7. The answer is the same for (b), and (c), see Problem 14.18.

15.19. See Chapter 24 for more general results.

16.5. Show that the set defined this way is open and closed.

16.6. With $\omega = df_\alpha$ on U_α, trivialize the covering over U_α, by mapping $p_\omega^{-1}(U_\alpha) \xrightarrow{\cong} U_\alpha \times \mathbb{R}$ by taking a germ f at P to $P \times (f(P) - f_\alpha(P))$. Check that $g_{\alpha\beta} = f_\alpha - f_\beta$ are transition functions for this covering. If $\omega' = \omega + dg$, let $f_\alpha' = f_\alpha + g$, and one obtains the same transition functions.

16.11. $H^0X \to H^0(X;\mathbb{R})$ comes from the fact that a locally constant function is a function on X with coboundary zero; $H^1X \to H^1(X;\mathbb{R})$ comes from the fact that a closed 1-form defines by integration a function on paths that is a 1-cocycle. It follows from Proposition 16.10 and Theorem 15.11 (and Exercise 15.18) that these maps are isomorphisms.

16.12. Use the ideas of Lemma 10.2. (Caution: Proposition 16.10 and Mayer–Vietoris for homology can be used directly for example when $G = \mathbb{R}$, but not in general.) See §24a for the general story.

16.13. Since the given covering is locally isomorphic to the trivial G-cov-

ering, it suffices to prove that p_T is a trivial covering when the given covering is the trivial G-covering. If the given covering is the projection from the product $X \times G \to X$, there is a canonical mapping

$$Y_T = ((X \times G) \times T)/G \to X \times T$$

determined by the map $\langle (x \times g) \times t \rangle \mapsto x \times g^{-1} \cdot t$. To see that this is well defined on orbits, note that, for h in G,

$$\begin{aligned} h \cdot \langle (x \times g) \times t \rangle &= \langle (x \times h \cdot g) \times h \cdot t \rangle \\ &\mapsto x \times (hg)^{-1} h \cdot t = x \times g^{-1} \cdot t. \end{aligned}$$

Check that this is a homeomorphism, with its inverse determined by sending $x \times t$ to $\langle (x \times e) \times t \rangle$.

16.14. For any y in Y and t in T, there is a unique element $\Phi(y, t)$ in T' so that

$$f(\langle y \times t \rangle) = \langle y \times \Phi(y, t) \rangle \qquad \text{for all } y \text{ in } Y \text{ and } t \text{ in } T.$$

For fixed t, the mapping $y \mapsto \Phi(y, t)$ from Y to T' is locally constant, since it is constant on each piece of $p^{-1}(N)$, for N an evenly covered set in X. But since Y is connected, a locally constant function is constant; therefore $\Phi(y, t) = \varphi(t)$ for some function $\varphi\colon T \to T'$. To see that φ is a map of G-sets, calculate:

$$\begin{aligned} \langle y \times \varphi(g \cdot t) \rangle &= f(\langle y \times g \cdot t \rangle) = f(\langle g^{-1} \cdot y \times t \rangle) \\ &= \langle g^{-1} \cdot y \times \varphi(t) \rangle = \langle y \times g \cdot \varphi(t) \rangle, \end{aligned}$$

from which the equation $\varphi(g \cdot t) = g \cdot \varphi(t)$ follows. By the definition, the mapping from Y_T to $Y_{T'}$ determined by φ is f.

16.17. There is a mapping from Y/H to $(Y \times G/H)/G$ that takes the H-orbit of a point y to the G-orbit of the point $y \times H$. The inverse mapping from $(Y \times G/H)/G \to Y/H$ is given by sending the G-orbit of $y \times gH$ to the H-orbit of $g^{-1} \cdot y$.

16.22. By Exercise 16.14, an isomorphism $f\colon Y(\psi_1) \to Y(\psi_2)$ is given by a map $\varphi\colon G' \to G'$ of G-sets. So

(i) $f(\langle z \times g' \rangle) = \langle z \times \varphi(g') \rangle$ for all $z \in Y$ and $g' \in G'$; and
(ii) $\varphi(g' \cdot \psi_1(g^{-1})) = \varphi(g') \cdot \psi_2(g^{-1})$ for all $g' \in G'$ and $g \in G$.

Since f preserves base points,

$$f(\langle y \times e \rangle) = \langle y \times \varphi(e) \rangle = \langle y \times e \rangle,$$

so we must have $\varphi(e) = e$. Since f is a mapping of G'-coverings, we must have

$$f(\langle z \times g' \rangle) = f(g' \cdot \langle z \times e \rangle) = g' \cdot f(\langle z \times e \rangle) = g' \cdot \langle z \times e \rangle = \langle z \times g' \rangle.$$

Therefore $\varphi(g') = g'$ for all g', and applying (ii) with $g' = e$ we see that $\psi_1(g^{-1}) = \psi_2(g^{-1})$ for all g in G, so $\psi_1 = \psi_2$.

16.24. Over U_α, where $p^{-1}(U_\alpha) \cong U_\alpha \times G$, identify $(p^{-1}(U_\alpha) \times T)/G$ with the product $U_\alpha \times T$ by $\langle x \times g \times t \rangle \mapsto x \times g^{-1} \cdot t$. The transition from $U_\alpha \times T$ to $U_\beta \times T$ is

$$x \times t \mapsto \langle x \times e \times t \rangle \mapsto \langle x \times g_{\alpha\beta}(x) \cdot e \times t \rangle \mapsto x \times g_{\alpha\beta}(x)^{-1} \cdot t = x \times g_{\beta\alpha}(x) \cdot t.$$

16.25. Consider first the case of a trivial covering $Y = X \times G$. Then

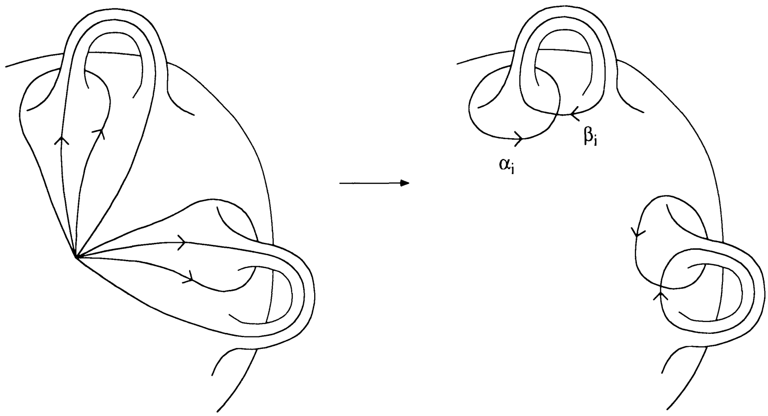
$$(Y \times G')/G = ((X \times G) \times G')/G \xrightarrow{\cong} X \times G', \quad x \times g \times g' \mapsto x \times g' \cdot \psi(g).$$

Choose trivializations $p^{-1}(U_\alpha) \cong U_\alpha \times G$ of the covering p, so that the resulting transitions are given by the cocycle $\{g_{\alpha\beta}\}$. Identify $p'^{-1}(U_\alpha)$ with $U_\alpha \times G'$ by the displayed isomorphism. The transition from $U_\alpha \times G'$ to $U_\beta \times G'$ (over $U_\alpha \cap U_\beta$) is

$$x \times g' \mapsto \langle x \times e \times g' \rangle \mapsto \langle x \times g_{\alpha\beta}(x) \cdot e \times g' \rangle \mapsto x \times g' \cdot \psi(g_{\alpha\beta}(x)).$$

16.28. The transition functions are given by the Jacobian determinants of the change of coordinate mappings.

17.10.



17.12. The fundamental group is the free group with $2g + n$ generators $a_1, b_1, \ldots, a_g, b_g, d_1, \ldots, d_n$, divided by the least normal subgroup containing

$$c_g = a_1 \cdot b_1 \cdot a_1^{-1} \cdot b_1^{-1} \cdot \ldots \cdot a_g \cdot b_g \cdot a_g^{-1} \cdot b_g^{-1} \cdot d_1 \cdot \ldots \cdot d_n.$$

The result is a free group on $2g + n - 1$ generators, if $n \geq 1$ (since one can write d_n in terms of the other generators).

18.3. These integers are determined by writing the class of γ in terms of the basis: $[\gamma] = \Sigma_{i=1}^{g}(m_i[a_i] + n_i[b_i])$.

18.4. If $\{\psi_\beta'\}$ is another partition of unity subordinate to another atlas of charts,

$$\sum_\beta \left(\iint_X \psi_\beta' \cdot \nu\right) = \sum_\beta \left(\sum_\alpha \left(\iint_X \psi_\alpha \cdot \psi_\beta' \cdot \nu\right)\right) = \sum_{\alpha,\beta} \left(\iint_X \psi_\alpha \cdot \psi_\beta' \cdot \nu\right)$$

$$= \sum_\alpha \left(\sum_\beta \left(\iint_X \psi_\beta' \cdot \psi_\alpha \cdot \nu\right)\right) = \sum_\alpha \left(\iint_X \psi_\alpha \cdot \nu\right).$$

The linearity is immediate from the definitions. For the opposite orientation, one can use the same charts but with the x and y axes interchanged, so in local coordinates the form ν is expressed as $-\nu_\alpha\, dy\, dx$, and all the integrals get replaced by their negatives.

18.7. Since the 1-forms α_i and β_i form a basis for H^1X, it suffices to prove these formulas when ω is one of the 1-forms α_i or β_i. Lemmas 18.1 and 18.6 imply these formulas. For example, (α_j, α_i) and $\int_{a_j}\alpha_i$ are both 0, as are (β_j, β_i) and $\int_{b_j}\beta_i$, and (α_j, β_i) and $\int_{a_j}\beta_i$ are both 1 if $i = j$, and 0 otherwise. Finally, $(\beta_j, \alpha_i) = -(\alpha_i, \beta_j)$, and $-(\alpha_i, \beta_j)$ and $\int_{b_j}\alpha_i$ are both -1 if $i = j$ and 0 otherwise.

18.8. By linearity, it is enough to do it for $\mu = \alpha_i$ and $\mu = \beta_i$, and it then follows from Exercise 18.7 and Lemma 18.1.

18.11. See Exercise 18.7 for (a).

18.12. Changing the a_i and b_i if necessary, one can assume they cross the annulus transversally. Calculate for $\mu = \alpha_i$ and β_i as above.

18.14. Use Problem 18.12.

18.15. It suffices to show that another choice of differentiable structure gives the same intersection numbers for pairs taken from basis elements a_i and b_j.

18.20. Use a partition of unity, as in Lemma 5.5.

18.22. If $H^2U = 0$ and $H^2V = 0$, it follows from this and the Mayer–Vietoris sequence in §16 that $H^2(U \cup V) = 0$. Use Exercise 18.18 and the fact that X can be built from rectangles, see Lemma 24.10.

18.25. One can realize the surface by removing h disks from a sphere, and gluing in h Moebius bands. Apply Mayer–Vietoris.

18.26. If f is a finite covering, and g is a function on Y, define $f_*(g)$ to be the function whose value at x is the sum of the values of g at the points of $f^{-1}(x)$. A similar definition works for forms.

19.1. $f'(z) = u_x + iv_x = v_y - iu_y$, and the Jacobian determinant is $u_xv_y - v_xu_y$.

19.4. Apply Riemann's theorem on removable singularities (Exercise 9.25).

19.6. For (d), multiply the function by a suitable $p_2(z)/p_1(z)$ so that the result has no poles, so is constant.

19.8. This follows from the fact that the map $z \mapsto z^e$ from D to D is proper, and the fact that only finitely many points are added.

19.17. (b) If $\lambda \cdot (\mathbb{Z} + \mathbb{Z}\tau') \subset \mathbb{Z} + \mathbb{Z}\tau$, $\lambda \cdot \tau' = a\tau + b$, $\lambda \cdot 1 = c\tau + d$; the determinant is ± 1 exactly when $\lambda \cdot (\mathbb{Z} + \mathbb{Z}\tau') = \mathbb{Z} + \mathbb{Z}\tau$, and it is positive if τ and τ' are both in the upper half plane.

20.5. To prove that C is a Riemann surface near such a point, show that one or the other projection to an axis is a local isomorphism.

20.6. The genus in each case is: (i) $(m-2)/2$ if m is even, $(m-1)/2$ if m is odd; (ii) 0; (iii) 4; (iv) 1; and (v) $(m-1)(m-2)/2$.

20.7. (c) With the notation of Exercise 19.12, the covering is given by assigning each σ_i to the unique transposition $(1\ 2)$.

20.12. The assumptions in (ii) guarantee that X has n distinct points over ∞, all therefore with ramification index 1. To see this, set $z' = 1/z$ and $w' = w/z$; they satisfy an equation $G(z', w') = \Sigma_{i=0}^{n} b_i(z')(w')^{n-1} = 0$, where $b_i(z') = (z')^i a_i(1/z')$ is a polynomial in z' with constant term λ_i; the n roots to the equation $\Sigma_{i=0}^{n} \lambda_i t^{n-i} = 0$ give n points on X over ∞.

Consider the zeros and poles of the meromorphic function h. To see what happens at the points over ∞, make the change of coordinates as above. A calculation shows that $h = (z')^{1-n} G_{w'}(z', w')$, so h must have pole of order $n-1$ at each of these n points. At the other points of X, i.e., the points of C, h has no poles, so by Corollary 19.5 the sum of the orders of zeros of h at the points of C must be $n(n-1)$. By Exercise 20.9, this sum is the sum of $e(P) - 1$ over all ramification points of the mapping $z\colon X \to S^2$. Apply Riemann–Hurwitz to prove (c).

The space of polynomials of degree at most $n-3$ in Z and W has a basis the monomials $Z^i W^j$ with $i + j \leq n - 3$, and the number of these is

$$(n-2) + (n-3) + \dots + 2 + 1 = (n-1)(n-2)/2.$$

If we verify that $g \cdot dz/h$ is a holomorphic 1-form for each $g = z^i w^j$, it follows that we have produced g_X holomorphic 1-forms, showing at once that these are all of the holomorphic 1-forms, and that their dimension is g_X. At the points in C, the form dz/h is holomorphic, as follows from the equation

$$0 = d(F(z,w)) = F_z(z,w)\,dz + F_w(z,w)\,dw,$$

so $dz/h = -dw/F_z(z,w)$, and one of the denominators is nonzero at each point of C. At infinity, since $dz = -(z')^{-2}dz'$,

$$\begin{aligned} z^i w^j \frac{dz}{F_w(z,w)} &= (z')^{-i}\left(\frac{w'}{z'}\right)^j \frac{-(z')^{-2}\,dz'}{(z')^{1-n}G_{w'}(z',w')} \\ &= -(z')^{n-3-i-j}(w')^j \frac{dz'}{G_{w'}(z',w')}, \end{aligned}$$

from which we see that these 1-forms are also holomorphic at points over ∞.

20.13. The proof is essentially the same, except that $F_w(z, w)$ vanishes at the 2δ points of X lying over the nodes, as well as at the branch points. Note that the space of polynomials of degree at most $n-3$ vanishing at δ points always has dimension at least $(n-1)(n-2)/2-\delta$, so the construction produces at least g_X independent holomorphic 1-forms. But since $\dim(\Omega^{1,0}) \leq g$, the inequalities are equalities—which shows, in fact, that conditions to vanish at the nodes are all independent.

20.16. Compare the sum of the orders of $f^*\omega$ with those of ω, for ω a meromorphic 1-form on Y.

20.17. See Exercise 9.26.

20.19. For (a), since the residue is linear, it is enough to prove it when $\omega = (z-a)^m\,dz$, $a \in \mathbb{C}$, $m \in \mathbb{Z}$. (b) can be reduced to a local calculation, over a disk in S^2, where one has explicit formulas for the map z.

20.24. The point is that, up to periods, integrating from P to Q and then from Q to R is the same as integrating from P to R.

21.2. Write $E = D + Q_1 + \ldots + Q_r$ and apply Lemma 21.1 r times.

21.13. Define the adele $\mathbf{f}$ to be f_i at P_i and 0 elsewhere. Take D of large degree so that $M + R(D) = R$, and with $\mathrm{ord}_{P_i}(D) \geq m_i$ for all i. There is therefore an f in M so that $f - \mathbf{f}$ is in $R(D)$.

21.20. For (a), if points $P_1, \ldots, P_k$ have been found so that $\dim(\Omega(P_1 + \ldots + P_k)) = g - k$, and $k < g$, take any nonzero ω in $\Omega(P_1 + \ldots + P_k)$, and let P_{k+1} be any point which is not a zero of ω. For (b), change the last P_g to be a zero of $\omega \neq 0$ in $\Omega(P_1 + \ldots + P_{g-1})$. For points as in (b), Riemann–Roch implies that $\dim(L(P_1 + \ldots + P_g)) \geq 2$, so there is a nonconstant function with at most g poles.

21.21. Multiplying φ by a scalar, we may arrange so the residues of φ at P and Q are as stated. This φ is unique up to adding a holomorphic ω. Take a_j, b_j, ω_j as in §20d, with the a_j and b_j not passing through P or Q. Use Corollary 20.22 to show that there are unique complex numbers λ_j so that the integral of $\varphi - \Sigma\lambda_j\omega_j$ over the cycles a_k and b_k are all purely imaginary.

21.22. Take φ as in the preceding exercise, and take u to be an integral of the real part of φ.

21.23. An element in $\Omega(-2P)$ that is not in $\Omega(0)$ must have a double pole, since it cannot have a residue at P. Multiply by a scalar to get $\varphi + dz/z^2$ holomorphic near P, change φ by a holomorphic 1-form to get all its periods purely imaginary, and integrate the real part of φ to get u.

21.24. Consider the sequence $\mathbb{C} = L(0) \subset L(P_1) \subset L(P_1 + P_2) \subset \ldots$, with

the dimensions going up by 0 or 1 at each step, reaching dimension g when $k = 2g - 1$.

21.25. Apply Riemann–Roch, with $E = K - D$.

21.27. The holomorphic differentials have the form $[(a + bz + cw)/w^3]\,dz$, with a, b, $c \in \mathbb{C}$. Check that no such form vanishes to order 2 at any point of X.

22.1. See the proof of Green's theorem in Appendix B. For the correct signs, see §23a.

22.2. Extend the definition of integral over continuous paths as in the plane, and use the same arguments as in the planar case.

22.3. 4π.

22.4. If $F\colon S^2 \to \mathbb{R}^3 \setminus \{0\}$, define $W(F, 0)$ to be $(1/4\pi) \iint_{F\circ\Gamma} \omega$.

22.6. Use Van Kampen, with one open set a ball around the missing point.

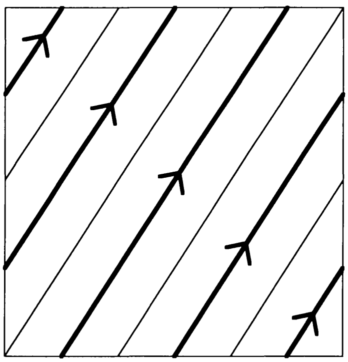
22.7. A homotopy is $H((z, w) \times s) = \left(sz, \sqrt{1 - s^2|z|^2}\,\dfrac{w}{|w|}\right)$.

22.9. To show the map is surjective, if (z, w) is in K, let $z' = -z/a$ and $w' = w/b$, and verify that $(z')^3 = (w')^2$ and $|z'| = |w'| = 1$.

22.11. For example, with $\rho(t) = 1 - t$, one may take

$$\lambda = {}^1\!/_2\left(-\rho^2 + \sqrt{4 + 4\rho^3 + \rho^4}\right) \quad \text{and} \quad \mu = \sqrt{\lambda - \rho}\,.$$

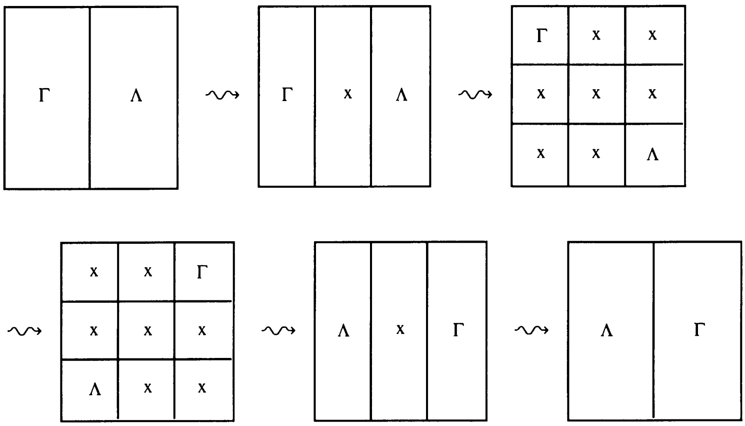
22.13. The generators for the fundamental groups of $A \setminus K$ and $B \setminus K$ are the circles around the middles of the solid tori. The generator for the fundamental group of $T \setminus K$ is a path as indicated:



To appeal to Van Kampen, A and B must be replaced by open neighborhoods of which they are retracts.

22.14. For (a), construct a mapping from $[0, 1]^k$ to S^k that maps the boundary to s_0, and is a homeomorphism from the interior of the cube to the complement of s_0. For (d), show that a map from S^k to S^n, with $k<n$, is homotopic to one that misses the south pole, say by triangulating S^k into small simplices, and approximating the map by a "simplicial" map. See, e.g., Hilton (1961).

22.16. Find homotopies that stretch and slide between the maps indicated in the diagram:



For example, the second homotopy maps $(t_1, \ldots, t_k, s)$ to

$$\begin{cases} \Gamma(3t_1, (3t_2 - 2s)/(3-2s), t_3, \ldots, t_k), & 0 \le t_1 \le 1/3, 2s/3 \le t_2 \le 1, \\ \Lambda(3t_1 - 1, 3t_2/(3-2s), t_3, \ldots, t_k), & 2/3 \le t_1 \le 1, 0 \le t_1 \le 1 - 2s/3, \\ x, & \text{otherwise}. \end{cases}$$

The third slides the squares around clockwise.

22.22. Details can be found in many texts, e.g., Bott and Tu (1980).

22.23. The form is closed by calculation, and it is not exact by Stokes theorem, since its integral over S^{n-1} is not zero.

22.27. See Bott and Tu (1980).

23.1. This is formal, using the identities

$$\partial_i^s \circ \partial_j^{s'} = \partial_{j-1}^{s'} \circ \partial_i^s \qquad \text{for} \quad i<j,$$

and s and s' each taking values 0 and 1. The signs cancel because of the shift in subscript from j to $j-1$.

23.5. Check that $(\partial_i^0 - \partial_i^1) \circ S = S \circ (\partial_i^0 - \partial_i^1)$.

23.15. Induct on p, and use Mayer–Vietoris for $U = U_1 \cup \ldots \cup U_{p-1}$ and

$V = U_p$. Note that $U \cap V = (U_1 \cap V) \cup \ldots \cup (U_{p-1} \cap V)$, so the inductive hypotheses imply to U, V, and $U \cap V$.

23.17. For (d), argue by induction, using Mayer–Vietoris. If $g\colon S^{n-1} \to S^{n-1}$ has degree d, show that $f\colon S^n \to S^n$ given by the formula

$$f(x_1, \ldots, x_{n+1}) = (g(x_1, \ldots, x_n), x_{n+1})$$

also has degree d.

23.19. For (c), use Problem C.3 and the n-dimensional version of Lemma B.9; see the proof of Claim D.3.

23.22. Induct on n, comparing the action of the antipodal map with the Mayer–Vietoris isomorphism.

23.23. For (a), such a vector field gives a map $f\colon S^n \to S^n$ that has no point P mapped to P or to $-P$. Such a map is homotopic to the identity and to the antipodal map. Use the preceding problem. For (b), consider the mapping $(x_1, x_2, \ldots, x_{n+1}) \mapsto (x_2, -x_1, \ldots, x_{n+1}, -x_n)$.

23.26. For (a), regard $S^m \subset S^n$, and note that a map $f\colon S^n \to S^n$ that is not surjective has degree zero. The other proofs are essentially the same as in Chapter 4.

23.27. In the situation of (a), there is a homotopy from f to a map g to which Theorem 23.24 applies, given by homotopic to the map g given by

$$H(P \times s) = \frac{f(P) - sf(P^*)}{\|f(P) - sf(P^*)\|}.$$

For (b) show similarly that f is homotopic to a map g with $g(P^*) = g(P)$ for all P. (c) Any automorphism g without fixed points has degree -1, so if g and h are nontrivial automorphisms, since $g \cdot h$ has degree 1, $g \cdot h$ must be the identity.

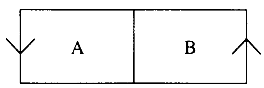
23.28. Lift to a map from S^n to S^n and apply Problem 23.27.

23.31. If $m = n$, $H_k(S^m \times S^n)$ is $\mathbb{Z}$ for $k = 0$ and $k = n + m$, and $\mathbb{Z} \oplus \mathbb{Z}$ if $k = n$, and 0 for other k. If $m \neq n$, the answer is $\mathbb{Z}$ for $k = 0$, m, n, and $m + n$, and 0 for other k. For the proof, induct on n, using $S^m \times U$ and $S^m \times V$, with U and V as before.

23.36. Identify $\mathbb{R}^n$ with the complement of a point P in S^n, and use Mayer–Vietoris for this open set and a small neighborhood of P.

23.37. See Proposition 5.17 and Corollary 5.18.

23.39. Cut the band in half:



. Look at the exact sequence

$$0 \to H_2(\mathbb{R}^3 \setminus A) \oplus H_2(\mathbb{R}^3 \setminus B) \to H_2(\mathbb{R}^3 \setminus A \cap B) \to H_1(\mathbb{R}^3 \setminus X) \to 0,$$

and the maps to the terms in the corresponding sequence with A and B replaced by $A \cap Y$ and $B \cap Y$, and X replaced by Y.

23.41. Use n-dimensional grids, and generalize the arguments of Chapters 6 and 9.

24.5. The fact that C'' is free abelian means that one can find a subgroup $\tilde{C}''$ of C that maps isomorphically onto C'', and then C is the direct sum of $\tilde{C}''$ and the image of C'; and $\mathrm{Hom}(-, G)$ preserves direct sums.

24.7. This amounts to an identity on I^k, which is proved just as in the case of Green's theorem for a rectangle by a calculation using Fubini's theorem, as in Proposition B.6.

24.9. This is the higher-dimensional version of reparametrization for paths. Here, for any k-cube Γ, $S \circ A(\Gamma) - \Gamma = \partial\Lambda$, where

$$\Lambda(s, t_1, \ldots, t_k) = \Gamma(s \cdot \alpha(2t_1) + (1 - s) \cdot t_1, \ldots, s \cdot \alpha(2t_k) + (1 - s) \cdot t_k),$$

noting that the other terms in the boundary are degenerate. Defining S_p and R_p by the same formulas as in §23b, one calculates that

$$S^p - 1 = \partial \circ R_p + R_p \circ \partial + (S \circ A - I) \circ S_p,$$

and the rest of the proof is the same as before.

24.14. This is clear except where the coboundary and dual of the boundary are involved. Let $\omega = \omega_1 - \omega_2$ be a closed $(k-1)$-form representing a class $[\omega]$ in $H^{k-1}(U \cap V)$, with ω_1 and ω_2 $(k-1)$-forms on U and V, and let $z = c_1 + c_2$ represent a class $[z]$ in $H_k^\infty(U \cup V)$, where c_1 and c_2 are chains on U and V respectively. Then

$$\delta([\omega])([z]) = \int_{c_1} d\omega_1 + \int_{c_2} d\omega_2 = \int_{\partial c_1} \omega_1 + \int_{\partial c_2} \omega_2 = [\omega](\partial([z])).$$

24.15. See Problem 23.41.

24.16. Use Mayer–Vietoris and induct on the number of open sets in the cover.

24.21. Set $(p_*\omega)(x) = \omega(y_1) + \omega(y_2)$ where $p^{-1}(x) = \{y_1, y_2\}$. For (c), note that $\int_{\tilde{X}} p^*\omega = 0$ for any n-form ω with compact support on X.

24.23. For (a), use $H^kX \cong (H^{n-k}X)^* \cong ((H^kX)^*)^*$, and the general fact that if V is a vector space isomorphic to $(V^*)^*$, then V must be finite dimensional. For (b), use the linear algebra fact that a finite-dimensional vector space with a nondegenerate skew-symmetric form must be even dimensional.

24.24. Since the spaces are finite dimensional, $H_c^{n-p}X$ is isomorphic to $(H^pX)^*$.

24.25. This follows from the fact that all cubes Γ have compact image in X. (Compare the special case used in the argument in §5c.)

24.26. (b) follows from the definitions, just as for homology.

24.27. Use Exercise 24.5.

24.36. Consider for each n the exact sequences $0 \to Z_n \to C_n \to B_{n-1} \to 0$ and $0 \to B_n \to Z_n \to H_n \to 0$, and use Problem C.16.

24.37. For (c), fix an ordering on $N(\mathfrak{U})$. For each vertex v, choose a point c_v in U_v, regarded as a 0-chain on U_v. If (v_0, v_1) is a 1-simplex, then, since $U_{v_0} \cup U_{v_1}$ is connected, there is a path from v_0 to v_1 in $U_{v_0} \cup U_{v_1}$, which determines a 1-chain $c_{(v_0,v_1)}$. Construct, by induction on n, for each n-simplex $(v_0, \ldots, v_n)$, a chain $c_{(v_0,\ldots,v_n)}$ in $C_n(U_{v_0} \cup \ldots \cup U_{v_n})$, such that

$$\partial(c_{(v_0,\ldots,v_n)}) = \sum_{i=0}^{n} (-1)^i c_{(v_0,\ldots,\hat{v}_i,\ldots,v_n)} .$$

The existence of $c_{(v_0,\ldots,v_n)}$ follows from the fact that the right side is an $(n-1)$-cycle on $U_{v_0} \cup \ldots \cup U_{v_n}$, and $H_{n-1}(U_{v_0} \cup \ldots \cup U_{v_n}) = 0$ by Exercise 23.15(b). This gives a map $C_*(N(\mathfrak{U})) \to C_*X$. To see that the resulting map on homology is an isomorphism, induct on the number of open sets; take any v and construct subcomplexes L_1 and L_2 as in the proof of Proposition 24.33, and compare the corresponding Mayer–Vietoris sequence for L_1 and L_2 with that of the covering of X by U_v and $\cup_{v' \neq v} U_{v'}$.

A.6. If not, each point has a neighborhood meeting only finitely many, and K would be contained in a finite union of such neighborhoods.

A.8. Without loss of generality, one may assume K contains a neighborhood of the origin. Map ∂K to S^{n-1} by mapping P to $P/\|P\|$. This is continuous and bijective, so a homeomorphism. Let $f\colon S^{n-1} \to \partial K$ be the inverse map. Define $F\colon D^n \to K$ by $F(0) = 0$, and $F(P) = \|P\| \cdot f(P/\|P\|)$ for $P \neq 0$. Then F is continuous and bijective, so a homeomorphism.

A.16. A connected and locally path-connected space is path-connected, as seen by showing that the set of points that can be connected to a given point by a path is open and closed.

B.12. Find a countable covering of U and so that each open set in the covering is contained in some U_α, and so that any point has a neighborhood that only meets finitely many of the open sets.

B.14. With g as in Step (2), let $h(x) = \int_0^x g(t)\,dt / \int_0^1 g(t)\,dt$, and set $\psi(x, y) = h(r - r_1)/r_2 - r_1)$, where $r = \|(x, y)\|$.

C.7. If $\{e_\alpha\}$ is a basis for F, and $\varphi(\tilde{e}_\alpha) = e_\alpha$, these $\tilde{e}_\alpha$ generate a free abelian subgroup $\tilde{F}$ of A, and A is the direct sum of $\mathrm{Ker}(\varphi)$ and $\tilde{F}$ by Exercise C.5.

C.15. Let α be the map from A to B, β the map from B to C, and let C' be the image of β, $\beta'\colon B \to C'$ the induced surjection. If f in $\mathrm{Hom}(B, \mathbb{R})$ maps

to 0 in Hom($A, \mathbb{R}$), f vanishes on the image of α. Since $B/\text{Image}(\alpha) \cong C'$, there is a homomorphism g' from C' to $\mathbb{R}$ such that $g' \circ \beta' = f$. By the lemma, there is a homomorphism g from C to $\mathbb{R}$ that restricts to g' on C'. Then $g \circ \beta = f$, which means that g in Hom($C, \mathbb{R}$) maps to f.

C.16. Let $B = A_2/A_1$, and show that there are exact sequences

$$0 \to A_1 \to A_2 \to B \to 0 \quad \text{and} \quad 0 \to B \to A_3 \to \ldots \to A_n \to 0.$$

References

Included in this list are books and articles that were referred to in the text. They can be consulted for more on topics covered in this book.

L.V. Ahlfors, *Complex Analysis*, 3rd edn, McGraw-Hill, 1979.

M.A. Armstrong, *Basic Topology*, Springer-Verlag, 1983.

R. Bott and L.W. Tu, *Differential Forms in Algebraic Topology*, Springer-Verlag, 1980.

G.E. Bredon, *Topology and Geometry*, Springer-Verlag, 1993.

S.S. Cairns, *Introductory Topology*, The Ronald Press, 1961.

C. Chevalley, *Introduction to the Theory of Algebraic Functions of One Variable*, American Mathematical Society, 1951.

W.G. Chinn and N.E. Steenrod, *First Concepts of Topology*, Random House, 1966.

H.S.M. Coxeter, *Introduction to Geometry*, 2nd edn, Wiley, 1961, 1980; Wiley Classics, 1989.

R.H. Crowell and R.H. Fox, *Introduction to Knot Theory*, Springer-Verlag, 1963.

J. Dieudonné, *A History of Algebraic and Differential Topology, 1990–1960*, Birkhäuser, 1989.

C. Godbillon, *Éléments de Topologie Algébrique*, Hermann, 1971.

M.J. Greenberg and J.R. Harper, *Algebraic Topology, A First Course*, Benjamin/Cummings, 1981.

P.A. Griffiths, *Introduction to Algebraic Curves*, American Mathematical Society, 1989.

V. Guillemin and A. Pollack, *Differential Topology*, Prentice-Hall, 1974.

D. Hilbert and S. Cohn-Vossen, *Geometry and the Imagination*, Chelsea, 1952.

P.J. Hilton, *An Introduction to Homotopy Theory*, Cambridge University Press, 1961.

J.G. Hocking and G.S. Young, *Topology*, Addison-Wesley, 1961, Dover, 1988.

H. Hopf, *Differential Geometry in the Large*, Springer Lecture Notes 1000, 1983.

F. Klein, *On Riemann's Theory of Algebraic Functions and their Integrals*, MacMillan and Bowes, 1893; Dover, 1963.

S. Lang, *Introduction to Algebraic and Abelian Functions*, 2nd edn, Springer-Verlag, 1982.

W.S. Massey, *A Basic Course in Algebraic Topology*, Springer-Verlag, 1991.

J. Milnor, *Morse Theory*, Annals of Mathematics Studies, Princeton University Press, 1963.

J. Milnor, *Topology from the Differentiable Viewpoint*, University Press of Virginia, 1965.

R. Narasimhan, *Compact Riemann Surfaces*, Birkhäuser, 1992.

M.H.A. Newman, *Elements of the Topology of Plane Sets of Points*, Cambridge University Press, 1939.

V.V. Nikulin and I.R. Shafarevich, *Geometries and Groups*, Springer-Verlag Universitext, 1987.

H. Rademacher and O. Toeplitz, *The Enjoyment of Mathematics*, Princeton University Press, 1957.

E.G. Rees, *Notes on Geometry*, Springer-Verlag, 1983.

G. Springer, *Introduction to Riemann Surfaces*, Addison-Wesley, 1957.

A. Wallace, *Differential Topology, First Steps*, W.A. Benjamin, 1968.

Index of Symbols

Index

Graduate Texts in Mathematics

(continued from page ii)

120 ZIEMER. Weakly Differentiable Functions: Sobolev Spaces and Functions of Bounded Variation.
121 LANG. Cyclotomic Fields I and II. Combined 2nd ed.
122 REMMERT. Theory of Complex Functions. *Readings in Mathematics*
123 EBBINGHAUS/HERMES et al. Numbers. *Readings in Mathematics*
124 DUBROVIN/FOMENKO/NOVIKOV. Modern Geometry—Methods and Applications. Part III.
125 BERENSTEIN/GAY. Complex Variables: An Introduction.
126 BOREL. Linear Algebraic Groups. 2nd ed.
127 MASSEY. A Basic Course in Algebraic Topology.
128 RAUCH. Partial Differential Equations.
129 FULTON/HARRIS. Representation Theory: A First Course. *Readings in Mathematics*
130 DODSON/POSTON. Tensor Geometry.
131 LAM. A First Course in Noncommutative Rings.
132 BEARDON. Iteration of Rational Functions.
133 HARRIS. Algebraic Geometry: A First Course.
134 ROMAN. Coding and Information Theory.
135 ROMAN. Advanced Linear Algebra.
136 ADKINS/WEINTRAUB. Algebra: An Approach via Module Theory.
137 AXLER/BOURDON/RAMEY. Harmonic Function Theory.
138 COHEN. A Course in Computational Algebraic Number Theory.
139 BREDON. Topology and Geometry.
140 AUBIN. Optima and Equilibria. An Introduction to Nonlinear Analysis.
141 BECKER/WEISPFENNING/KREDEL. Gröbner Bases. A Computational Approach to Commutative Algebra.
142 LANG. Real and Functional Analysis. 3rd ed.
143 DOOB. Measure Theory.
144 DENNIS/FARB. Noncommutative Algebra.
145 VICK. Homology Theory. An Introduction to Algebraic Topology. 2nd ed.
146 BRIDGES. Computability: A Mathematical Sketchbook.
147 ROSENBERG. Algebraic *K*-Theory and Its Applications.
148 ROTMAN. An Introduction to the Theory of Groups. 4th ed.
149 RATCLIFFE. Foundations of Hyperbolic Manifolds.
150 EISENBUD. Commutative Algebra with a View Toward Algebraic Geometry.
151 SILVERMAN. Advanced Topics in the Arithmetic of Elliptic Curves.
152 ZIEGLER. Lectures on Polytopes.
153 FULTON. Algebraic Topology: A First Course.
154 BROWN/PEARCY. An Introduction to Analysis.
155 KASSEL. Quantum Groups.
156 KECHRIS. Classical Descriptive Set Theory.
157 MALLIAVIN. Integration and Probability.
158 ROMAN. Field Theory.
159 CONWAY. Functions of One Complex Variable II.
160 LANG. Differential and Riemannian Manifolds.
161 BORWEIN/ERDÉLYI. Polynomials and Polynomial Inequalities.
162 ALPERIN/BELL. Groups and Representations.
163 DIXON/MORTIMER. Permutation Groups.
164 NATHANSON. Additive Number Theory: The Classical Bases.
165 NATHANSON. Additive Number Theory: Inverse Problems and the Geometry of Sumsets.
166 SHARPE. Differential Geometry: Cartan's Generalization of Klein's Erlangen Program.
167 MORANDI. Field and Galois Theory.
168 EWALD. Combinatorial Convexity and Algebraic Geometry.
169 BHATIA. Matrix Analysis.
170 BREDON. Sheaf Theory. 2nd ed.
171 PETERSEN. Riemannian Geometry.
172 REMMERT. Classical Topics in Complex Function Theory.
173 DIESTEL. Graph Theory. 2nd ed.
174 BRIDGES. Foundations of Real and Abstract Analysis.
175 LICKORISH. An Introduction to Knot Theory.
176 LEE. Riemannian Manifolds.
177 NEWMAN. Analytic Number Theory.
178 CLARKE/LEDYAEV/STERN/WOLENSKI. Nonsmooth Analysis and Control Theory.
179 DOUGLAS. Banach Algebra Techniques in Operator Theory. 2nd ed.

180 SRIVASTAVA. A Course on Borel Sets.
181 KRESS. Numerical Analysis.
182 WALTER. Ordinary Differential Equations.
183 MEGGINSON. An Introduction to Banach Space Theory.
184 BOLLOBAS. Modern Graph Theory.
185 COX/LITTLE/O'SHEA. Using Algebraic Geometry.
186 RAMAKRISHNAN/VALENZA. Fourier Analysis on Number Fields.
187 HARRIS/MORRISON. Moduli of Curves.
188 GOLDBLATT. Lectures on the Hyperreals: An Introduction to Nonstandard Analysis.
189 LAM. Lectures on Modules and Rings.
190 ESMONDE/MURTY. Problems in Algebraic Number Theory.
191 LANG. Fundamentals of Differential Geometry.
192 HIRSCH/LACOMBE. Elements of Functional Analysis.
193 COHEN. Advanced Topics in Computational Number Theory.
194 ENGEL/NAGEL. One-Parameter Semigroups for Linear Evolution Equations.
195 NATHANSON. Elementary Methods in Number Theory.
196 OSBORNE. Basic Homological Algebra.
197 EISENBUD/ HARRIS. The Geometry of Schemes.
198 ROBERT. A Course in p-adic Analysis.
199 HEDENMALM/KORENBLUM/ZHU. Theory of Bergman Spaces.
200 BAO/CHERN/SHEN. An Introduction to Riemann–Finsler Geometry.
201 HINDRY/SILVERMAN. Diophantine Geometry.